The Dissertation Committee for Andrea Lynn Godfrey
certifies that this is the approved version of the following dissertation:

A Product Segmentation Approach and its Relationship to Customer Segmentation Approaches and Recommendation System Approaches

Committee:

Leigh McAlister, Supervisor

Frenkel ter Hofstede

Romana Khan

Maytal Saar-Tsechansky

Rajashiri Srinivasan

A Product Segmentation Approach and its Relationship to Customer Segmentation Approaches and Recommendation System Approaches

by

Andrea Lynn Godfrey, B.Eng., M.B.A.

Dissertation

Presented to the Faculty of the Graduate School of

the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2007

DEDICATION

I dedicate this dissertation to my family. To my parents, Grant and Beverly Godfrey, for providing me with every opportunity I desired throughout my life and, in particular, for their love and support in my decision to pursue this doctoral degree. And to my sister, Trisha Fitzgerald, for always being my greatest cheerleader and closest confidante. I could never adequately express how fortunate I am to have such a wonderful family.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my advisor, Leigh McAlister, for providing me with outstanding training throughout my doctoral studies and for serving as such an excellent model of a passionate and generous researcher and teacher. I would like to thank the members of my dissertation committee, Frenkel ter Hofstede, Romana Khan, Maytal Saar-Tsechansky, and Rajashiri Srinivasan for their valuable input in this research. I am extremely fortunate to have had the opportunity to work with each of you. Special thanks go to Maytal Saar-Tsechansky for being a wonderful colleague and role model. I would also like to thank my former professor, Kathleen Seiders, for directing me toward pursing doctoral studies and for her continued friendship and guidance. I thank Rafael Bracero for all of his encouragement and for sharing my enthusiasm for marketing. Finally, I thank my family for their continued love and support.

A Product Segmentation Approach and its Relationship to

Customer Segmentation Approaches and

Recommendation System Approaches

Publication No. __________

Andrea Lynn Godfrey, Ph.D.
The University of Texas at Austin, 2007

Supervisor: Leigh McAlister

As part of their customer management strategy, retailers with large, multi-category offerings need to present their products in ways that help target customers search and choose from those offerings. In the first essay of this dissertation, a product segmentation approach is proposed. The proposed approach gives retailers a methodology for directly identifying customer-centric, cross-category, product segments from large numbers of products in multiple categories such that products within a segment are purchased by the same type of customers. In addition, the research examines the relationship between the proposed product segmentation approach and a parallel customer segmentation approach. The close relationship between the approaches suggests that the segments of products and customers inferred from each approach will be equivalent. However, the results show that this is not the case because of the aggregation constraint imposed on customers in the product segmentation approach and on products in the

customer segmentation approach. Further, the results indicate that the product segmentation approach provides better recommendations of products for a customer to purchase, while the customer segmentation approach provides better recommendations of customers for a product to target.

To increase customer repurchasing and loyalty, retailers with large offerings are increasingly employing recommendation systems. The second essay of this dissertation contributes to our understanding of recommendation systems in three respects. First, we present a new methodology, attribute-based co-clustering, which incorporates customer characteristics and product attributes to produce recommendations. This approach has not previously been evaluated in recommendation system contexts. Second, we compare the performance of the proposed approach with a related latent class segmentation approach and a widely applied collaborative filtering approach. Third, we identify factors that impact recommendation quality in two contexts: recommending a set of products for a customer to purchase and recommending a set of customers for a product to target. Results indicate that latent class segmentation quality improves in databases with large samples and strong predictors that characterize customers' preferences, while collaborative filtering quality improves with greater data density. Attribute-based co-clustering quality improves when customer and product attributes are predictive of choices, however, it is more stable with respect to data distribution.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ESSAY 1: IDENTIFYING CUSTOMER-CENTRIC, CROSS-CATEGORY PRODUCT GROUPS: A PRODUCT SEGMENTATION APPROACH AND ITS RELATIONSHIP TO CUSTOMER SEGMENTATION APPROACHES

CHAPTER 1: INTRODUCTION

Many retailers present their customers very large, multi-category product or service offerings. Category specialist retailers such as Best Buy, Bed Bath and Beyond, and Staples offer 20,000 to 40,000 SKUs in their stores (Levy and Weitz 2006), while a typical Wal-Mart supercenter offers over 150,000 SKUs (Daniels 2004). As part of their customer management strategies, retailers that offer such large numbers of products or services need to align their offerings with the types of customers they wish to target. That is, the retailers need to present their product or service offerings in ways that will help their target customers efficiently search and choose products or services from the large offerings. (Henceforth, we use the term "products" to refer equally to products or services.)

The following example illustrates how a retailer with a large, multi-category product offering might attempt to align its offering and its target types of customers. Wal-Mart sells thousands of products in a number of different categories including household products, personal care items, apparel, and groceries. Wal-Mart recently identified six types of customers it wishes to target: "Hispanics," "African-Americans," "Suburbanites," "Rural Residents," "Affluent," and "Empty-nesters" (McTaggert 2006). Currently, the retailer's store layouts are organized by category. However, to improve the in-store shopping experience of its six target types of customers, Wal-Mart could organize its store layouts into collections of products that attract each of the target customer-types. Similarly, to help its customers better navigate its online retail channel, Wal-Mart could structure its website format such that web pages feature collections of products that each customer-type is most likely to find appealing. To develop the store layouts or website format, the retailer needs to identify the subset of its products that is most attractive to each target customer-type.

In general, to present their offerings in a way that is more appealing to their target customers, retailers with large, multi-category product offerings require a methodology for identifying customer-centric, cross-category groups of products. We refer to these product groupings as "product segments". This terminology is commonly used in marketing practice to refer to groups of products that attract a particular customer-type. In the packaged goods industry, product segments are sometimes named by the benefit that attracts customers, e.g., low-carbohydrate, low-fat, organic, and sugar-free product segments (Convenience Store News 2004). In the automobile industry, product segments are sometimes named by the demographic that characterizes attracted customers, e.g., young driver or family car segments (Economist 2005). However, such product segments are defined based on managerial intuition into which products should be included in a segment rather than empirical analysis of the type of customers attracted to the collection of products. In this research, we present a methodology for defining product segments by empirically identifying groups of products from multiple categories such that products in a group attract the same type of customers.

One approach that marketers have traditionally used to align customers and products is customer segmentation, and one of the most popular approaches to segmenting customers is to combine latent class analysis with a choice model (e.g., Heilman and Bowman 2002; Kamakura and Russell 1989). Latent class customer segmentation aligns customers and products by identifying groups of customers such that customers within a group prefer the same type of products. However, this approach may be limiting in two respects. First, the approach may be limiting for retailers that offer very large numbers of products because it requires the analyst to aggregate the many individual products into a smaller number of choice alternatives to make estimation and interpretation of the underlying choice model tractable. This aggregation is based on the analysts' assumption of the product attribute that drives customers' preferences and varies across applications. For example, Kamakura and Russell (1989) aggregated the SKUs in a single consumer packaged food category into four alternatives based on brand name: national brands A, B, and C, and a composite brand, P, representing private label and regional brands. Heilman and Bowman (2002) aggregated 40 SKUs in three categories of baby products – disposable diapers, formula and towels – into 20 alternatives based on market share: big/leading national brands, medium size brands, small brands, private label composites, and "other" brand composites. While aggregating choice alternatives is standard procedure in such models, this aggregation imposes a constraint on the analysis. That is, in aggregating SKUs, the analyst assumes that customers' preferences for each of the aggregated SKUs is identical and is equal to the estimated preference for the aggregate product. As such, the aggregation constraint can potentially obscure information and restrict the analyst's ability to understand the relationship between customers and products at the most disaggregated product level, such as a SKU.

Latent class customer segmentation may also have limitations in contexts where the retailer has defined the types of customers it wishes to target. Latent class customer segmentation first specifies managerially relevant types of products and then identifies, for each product-type, latent groups of customers that have the strongest preference for that product-type.

For example, Heilman and Bowman (2002) define 20 product-types as described above and then identify latent customer segments based on individual customers' preferences for those product-types. If the latent customer segments identified do not correspond to the retailer-defined target customer-types, the customer segmentation approach could be limited in the extent to which its results are actionable.

Another approach marketers use to understand the relationship between customers and products is market structure analysis. One can think of the market structure literature as providing models that identify product segments; the goal typically stated for such models is to group together products that a consumer would be willing to substitute for one another. By definition, then, a group of products indirectly inferred from market structure analysis would be composed of products that tend to attract the same type of customers. Traditionally, however, market structure models tend to consider only those products within a single category, such as coffee (e.g., Cooper 1988; Fraser and Bradford 1983), soft drinks (e.g., DeSarbo and De Soete 1984; Rao and Sabavala 1981), and laundry detergent (e.g., Elrod and Keane 1995; Ramaswamy and DeSarbo 1990). While some market structure analyses do consider multiple categories, they typically evaluate preferences and identify structure in each category independently (e.g., Erdem 1996; Grover and Dillon 1985; Russell and Bolton 1988; Shugan 1987). Erdem and Winer (1999) estimate consumer preferences in two closely related categories (toothpaste and toothbrushes) comprising substitute and complementary products by allowing price and attribute preferences to be correlated across categories. However, Erdem and Winer (1999) focus on mapping competitive relationships among brands in each category separately. In general, market structure models have traditionally not addressed contexts that involve identifying groups of products with associations across categories.

The first objective of this research is to present a methodology for identifying latent product segments from large numbers of products in multiple categories. The proposed method adapts the widely applied latent class methodology for identifying customer segments. While the latent class approach is used in customer segmentation to identify groups of *customers*, such that customers within a group prefer the same type of products, we use a latent class approach to identify groups of *products* such that products within a group attract the same type of customers. The proposed product segmentation approach identifies these product groupings by first defining a set of managerially relevant "customer-types". For example, for Wal-Mart, those managerially relevant customer-types are "Empty-nesters," "Affluent," etc. McAlister, George and Chien (2007) develop the attraction model, which allows one to determine, for a given product, the strength with which that product attracts the defined customer-types. Our approach combines latent class analysis with the attraction model to identify groups of products such that products in an identified group attract the same customer-types. Thus, our first contribution is to provide retailers and analysts with an empirical methodology for directly identifying cross-category, customer-centric product segments from large, multi-category product offerings.

The second objective of this research is to investigate the relationship between product segmentation and customer segmentation. We examine the relationship between the two approaches by empirically comparing the proposed use of latent class analysis to identify product segments with analogous use of latent class analysis to identify customer segments. First, we compare the two approaches in terms of the product segments and customer segments identified by each approach when applied to the same customer product choice data set. In an illustrative application involving an education service provider, we show that, contrary to prior suggestions (see Grover and Srinivasan 1987), the product segments and customer segments identified by one approach are not identical to the product segments and customer segments identified by the other approach. This happens because each approach is impacted by the aggregation constraints imposed on customers in the product segmentation approach and imposed on products in the customer segmentation approach. Thus, our second contribution is to enhance our understanding of the relationship between customer segmentation and product segmentation by illustrating the impact of the aggregation constraint on the underlying models.

In addition, we examine the relationship between product segmentation and customer segmentation by empirically comparing each approach's ability to address two questions of managerial relevance: Which group of products would one recommend for a customer to purchase? Which group of customers would one recommend for a product to target? The results of our illustrative application indicate that the product segmentation approach is more effective at recommending a group of products for a customer to purchase, while the customer segmentation approach is more effective at recommending a group of customers for a product to target. As such, our third contribution is to suggest managerial applications for which the proposed product segmentation approach and the traditional customer segmentation approach are likely to be relatively more effective.

In the sections that follow, we describe the relationship between the proposed latent class product segmentation approach and the widely applied latent class customer segmentation approach. We develop the model underlying the product segmentation approach and contrast the elements of that model with those of the parallel customer segmentation approach. In a service provider context, we estimate the two models and compare the relative efficacy of each approach's recommendations of (1) a group of products for a customer to purchase and (2) a group of customers for a product to target. We conclude with research and managerial implications and directions for future research.

CHAPTER 2: RELATIONSHIP BETWEEN PRODUCT SEGMENTATION AND CUSTOMER SEGMENTATION

In simultaneously estimating latent customer segments and market structure (or what we refer to as indirectly inferred product segments), Grover and Srinivasan (1987, p.140) suggest that, when brand choice probabilities are used as the basis for segmentation, the two analyses are "reverse sides of the same analysis." That is, customer segmentation, which directly identifies groups of customers, also implies product segments when one observes the products preferred by the different groups of customers. Similarly, then, product segmentation, which directly identifies groups of products, also implies customer segments when one observes the customers attracted to the different groups of products. Despite the parallel nature of the two analyses, however, it is not the case that they yield exactly the same results when applied to a given data set. In fact, the customer segments directly identified from a particular set of customer product choice data need not be identical to the customer segments inferred by product segmentation of the same data. Similarly, the product segments directly identified from a particular set of customer product choice data need not be identical to the product segments inferred by customer segmentation of the same data.

Directly identified product segments will differ from the product segments inferred by customer segmentation approaches because of the aggregation constraint applied to the individual choice alternatives when estimating the choice model underlying the customer segmentation approach. Because there typically are a large number of individual choice alternatives in a raw consumer choice data set, such as a customer management database or scanner panel, analysts group those alternatives into managerially relevant aggregates and then estimate choice model parameters by considering the strength with which customers prefer those aggregates. To simplify exposition, we refer to individual choice alternatives as "products", and refer to the analyst-imposed, managerially relevant aggregates of those individual choice alternatives as "product-types." For example, we refer to the "big national brand" and "private label composite" defined by Heilman and Bowman (2002) as product-types. While the constraint of aggregating products into product-types makes choice model estimation feasible, the aggregation constraint also obscures some information about customers' preferences for the underlying products. For example, Heilman and Bowman's (2002) assumption that a customer will equally prefer all "private label composite" diapers may obscure heterogeneity in the degree to which that customer prefers individual SKUs defined as "private label composite".

Just as the choice model underlying the customer segmentation approach requires imposing an aggregation constraint on products, the attraction model (McAlister, George and Chien 2007) underlying our product segmentation approach requires imposing an aggregation constraint on customers. Because the number of individual customers in a consumer choice data set can be large, we first group individual customers into managerially relevant aggregates and then estimate attraction model parameters by considering the strength with which products attract those aggregates. To simplify exposition, we refer to individual customers as “customers”, and refer to the analyst-imposed, managerially relevant aggregates as “customer-types.” For example, we refer to the “Affluent” and “Empty-nesters” targeted by Wal-Mart as customer-types. While the constraint of aggregating customers into customer-types makes attraction model estimation feasible, the aggregation constraint also obscures some information about the strength with which different products attract individual customers. For example, Wal-Mart’s assumption that a product will equally attract all “Empty-nesters” would obscure heterogeneity in the degree to which that product attracts individual customers defined as “Empty-nesters”.

Because some information is obscured when imposing the product-type aggregation constraint in the customer segmentation approach, we expect this will adversely affect the quality of the product segments inferred from this approach. Similarly, because some information is obscured when imposing the customer-type aggregation constraint in the product segmentation approach, we expect this will adversely affect the quality of the customer segments inferred from this approach. Managerially, the quality of the results will impact each approach’s applicability in addressing different marketing problems. That is, we expect that the customer segmentation approach should be more effective than the product segmentation approach at recommending a group of customers for a product to target. Conversely, we expect that the proposed product segmentation approach should be more effective than the customer segmentation approach at recommending a group of products for a customer to purchase.

In summary, our proposed product segmentation approach and the parallel customer segmentation approach yield related results. However, because of the aggregation constraint imposed on customer data when applying the product segmentation approach and the aggregation constraint imposed on product data when applying the customer segmentation approach, one should not expect to identify the same product groupings and customer groupings when the two approaches are applied to the same data set. Further, it is likely that each approach will be better suited to addressing different managerial applications.

CHAPTER 3: MODEL DEVELOPMENT

Proposed Product Segmentation Approach

The proposed product segmentation approach begins with data for a sample of customers' choices across many products in multiple categories. The problem is made tractable by reducing the large number of individual customers to a few managerially relevant customer-types. For each product, we determine the relative strength with which that product attracts the different customer-types. We then identify a finite number of latent product segments such that products in a segment attract the same customer-types. Finally, in a posterior analysis, we probabilistically assign products to segments.

Determine the strength with which a product attracts different customer-types. To determine the relative strength with which a product attracts different customer-types, we apply McAlister, George and Chien's (2007) attraction model. This model represents the strength with which a product attracts each customer-type as the conditional probability that a given purchase of the product, p, was made by a particular customer-type, C, on a particular transaction, t, $prob(C,t \mid p)$. Specifically, given a set of customer-types, C =1,2,…, N_C and a set of individual products, p = 1,2,…,n_P, we define, for the randomly selected t^{th} transaction on which product p was chosen, the probability that the purchase was made by a customer of type C as:

$$prob(C,t \mid p) = \exp\{\alpha_{p,C,t}\} \Big/ \sum_{C=1}^{N_C} \exp\{\alpha_{p,C,t}\} \tag{1}$$

Given a set of M observable customer, product, and market environment variables that influence the strength with which product p attracts different customer-types, the deterministic component of the strength with which product p attracts a customer of type C on the t^{th} transaction is calculated as:

$$\alpha_{p,C,t} = \sum_{m=1}^{M} w_m x_{p,C,t} \tag{2}$$

where $x_{p,C,t}$ is the observed value of characteristic m for customer-type C and product p on the t^{th} transaction, and w_m is the attraction weight of characteristic m. The calculated probabilities represent a product's probability of attracting each customer-type on the t^{th} transaction. We refer to this set of probabilities as a product's "customer mix". The objective, assumptions, and specification of the attraction model are further outlined in Appendix A.

Identify product segments. We define product segments by identifying products that have the same customer mix. That is, we allow for heterogeneity in the strength with which products attract different customer-types. We group together products using a mixture model that combines latent class analysis with the attraction model. Specifically, we assume there exists a finite number of product segments N_Π and define, for any given purchase of product p in product segment Π, the probability that the purchase was made by customer-type C on the t^{th} transaction as:

(3) $$prob(C,t \mid p\varepsilon\Pi) = Q_\Pi * prob(C,t \mid p)$$

where $Q_\Pi = \exp\{\theta_\Pi\} \Big/ \sum_{\Pi=1}^{N_\Pi} \exp\{\theta_\Pi\}$ is the unconditional probability that a given product p is included in product segment Π, and θ_Π is the estimated product segment size parameter. We refer to this as the product segmentation (PS) approach. We estimate the model using maximum likelihood procedures to obtain estimates of the attraction weights for each product segment and the size of each product segment. Letting H_p be the collection of all transactions in which product p was chosen, the likelihood function is:

(4) $$L(H_p) = \sum_{\Pi=1}^{N_\Pi} Q_\Pi * L(H_p \mid \Pi)$$

where $L(H_p \mid \Pi) = \prod_{C=1}^{N_C} \prod_{t=1}^{T} prob(C,t \mid p\varepsilon\Pi)$.

Assign products to product segments. Finally, in a posterior analysis, we probabilistically assign each product to a product segment such that items assigned to a product segment attract the same customer-types. Specifically, we employ a Bayesian calculation to compute the probability that product p is included in product segment Π and assign each product to the product segment for which it has the highest inclusion probability. The segment assignment probabilities are calculated as:

(5) $$prob(p\varepsilon\Pi \mid H_p) = L(H_p \mid \Pi) * Q_\Pi \Big/ \sum_{\Pi=1}^{N_\Pi} [L(H_p \mid \Pi) * Q_\Pi]$$

Parallel Customer Segmentation Approach

The proposed product segmentation approach parallels latent class approaches that identify customer segments such that customers in a segment prefer the same product-types. In particular, the approach most closely parallels the approach first presented by Kamakura and Russell (1989) and recently presented by Heilman and Bowman (2002) to identify customer segments based on customers' preferences for product-types in multiple categories. We refer to this as the customer segmentation (CS) approach. In multi-category contexts, the CS approach begins with data for a sample of customers' choices across many products in multiple categories. In this case, the problem is made tractable by reducing the large number of products to a few managerially relevant product-types. Using a multinomial logit choice model, for each customer, the CS approach determines the relative strength with which that customer prefers the different product-types. Using latent class analysis, the CS approach identifies a finite number of latent customer segments and, in a posterior analysis, probabilistically assigns customers to segments. We compare in further detail the specification and estimation of the proposed PS approach and the CS approach in Appendix A.

Comparing Approaches

Given that the PS approach and the CS approach are parallel methodologies for aligning products and customers, we compare the approaches by testing their effectiveness in two managerial applications. The first test investigates each approach's effectiveness at recommending a group of products for a customer to purchase. Specifically, we compare the set of products each approach recommends for a withheld customer to purchase with the set of products that customer actually purchased. The second test investigates each approach's effectiveness at recommending a group of customers for a product to target. In this case, we compare the set of customers each approach recommends for a withheld product to target with the set of customers who actually purchased that product. In both tests, for each approach, we test the success of the approach using the leave-one-out variation of the n-fold bootstrapping technique (see Mitchell 1997). In a data set with N observations, this technique involves estimating the model N separate times on all of the data except for one observation (i.e., estimate the model with N-1 observations) and then making a prediction for the withheld observation.

First test: Recommend products for a customer to purchase. In the first test, we apply the PS approach and CS approach to recommend a group of products for a customer to purchase. For a withheld customer, we disregard all information about the customer except the customer-type, estimate the model using the purchase histories of all other customers in the data set, and then identify a group of products to recommend to the withheld customer. We repeat the process for each customer in the data set and calculate the hit rate of our recommendations (i.e., the extent to which the recommended products were actually purchased by the withheld customer) for each customer.

In the first test, recommendations of a group of products for a particular customer-type, C_0, to purchase are based on conditional probabilities $prob(\Pi \mid C_0)$ for the PS approach and $prob(P \mid C_0)$ for the CS approach (calculation of the conditional probabilities is presented in the top half of Appendix B). Both quantities report the probability that a particular kind of product was chosen (product from segment Π for the PS approach and product of type P for the CS approach) given that a customer of type C_0 did the choosing. Note, however, that for the PS approach, latent class analysis identifies the products to include in product segment Π, while for the CS approach, the products included in product-type P are determined a priori by the analyst. As such, we expect that the PS approach, which groups products using latent class analysis rather than analyst judgment, will be more effective at identifying a group of products to recommend for a customer to purchase.

Second test: Recommend customers for a product to target. In the second test, we apply the PS approach and the CS approach to recommend a group of customers for a product to target. For a withheld product, we disregard all information about the product except its product-type, P_0, estimate the model using the purchases of all other products in the data set, and then identify a group of customers to whom the withheld product should be targeted. We repeat the process for each product in the data set and calculate the hit rate of our recommendations (i.e., the extent to which the recommended customers actually purchased the product) for each product.

In the second test, recommendations of a group of customers for a particular product-type to target are based on conditional probabilities: $prob(C \mid P_0)$ for the PS approach and $prob(\chi \mid P_0)$ for the CS approach (calculation of the conditional probabilities is presented in the bottom half of Appendix B). Both quantities report the probability that a particular kind of customer did the choosing (customer of type C for the PS approach and a customer from segment χ for the CS approach) given that a product of type P_0 was chosen. Note, however, that for the CS approach, latent class analysis identifies the customers to include in customer segment χ, while for the PS approach the customers included in customer-type C are determined a priori by the analyst. As such, we expect that the CS approach, which groups customers using latent class analysis rather than analyst judgment, will be more effective at identifying a group of customers to recommend for a product to target.

Converting conditional probabilities into recommendations. We considered four rules for converting the conditional probabilities into recommendations. For ease of exposition, we discuss only one of these rules and present the additional rules and the results of their related statistical tests in Appendices C and D. The rule for which we report results recommends a product or customer if the conditional probability is greater than one would expect randomly. Specifically, for the first test, this rule recommends a group of products to a customer of type C_0 if the conditional probability of selecting a product from the group is greater than one would expect randomly (i.e., > [1 / number of sets of products]). In this test, we define the "hit rate" as the proportion of products the withheld customer actually purchased that the approach recommended. For the second test, this rule recommends a group of customers for a product of type P_0 to target if the conditional probability of selecting a customer from the group is greater than one would expect randomly (i.e., > [1 / number of sets of customers]). In this test we define the "hit rate" as the proportion of customers who actually purchased the withheld product that the approach recommended as targets.

CHAPTER 4: ILLUSTRATIVE APPLICATION

To illustrate the PS approach and compare the relative efficacy of the PS and CS approaches for different managerial objectives, we apply both approaches in a service context. Specifically, we examine the elective courses chosen by MBA students in the business school at a large southwestern university. As a service provider, the business school offers a large number of elective courses to meet the needs of different types of students. Specifically, the business school offers a range of elective courses across multiple departments (accounting, finance, management, management of information systems, and marketing) to meet the needs of students obtaining an MBA degree to pursue careers in investment banking, corporate finance, technology management, general management, brand management, and consulting, among other fields. As such, the business school is a service provider for which the products (services) are courses and the product-types (service-types) can be defined by the departments that offer those courses; and for which the customers are students and the customer-types can be defined by the careers that students want to pursue.

The problem of aligning the elective courses offered by the business school with the needs of different types of students can be viewed from two perspectives. First, consistent with the PS approach, one can consider the different courses that attract a particular student-type to determine the set of courses that could be recommended to that student-type. Understanding the set of courses that attracts a certain student-type can help the business school in a number of student management activities including preparing promotional materials to recruit students and recommending elective courses consistent with students' job placement objectives. Alternatively, consistent with the CS approach, one can consider the different students who prefer a particular course-type to determine the set of students to whom that course-type could be targeted. Understanding the set of students that prefers a certain course-type allows the business school to target a particular course to those students most likely to be interested in that course and to tailor a course's content to the goals of the students attracted to the course.

Description of Data

The data used to estimate the models includes two sets of information. The first data set describes the products (courses) and comprises course enrollment data for 32 elective MBA courses offered during the 1998-99 and 1999-2000 academic years as reported by the university's MBA program office. Compulsory courses were omitted from the analysis because these courses are required of all students and, therefore, have no observable variation in attraction across students. We find variation among elective course choices because students in this MBA program are not required to declare a concentration, but rather can choose courses offered by any department based on their interests, strengths and perspectives on how best to prepare for a particular career. Based on input from MBA program administrators on factors that might help explain courses' attraction for different students, the product attributes included in the analysis were the department in which the course is offered and the average course evaluation score. Courses are offered by five different departments: (1) Accounting, (2) Finance, (3) Management of Information Systems, (4) Management, and (5) Marketing. For estimation of the CS approach, we define the product-types by the department in which the course is offered. A summary of the product characteristics is presented in Table 1.

The second data set describes the sample of 326 customers (students) who graduated from the MBA program in 2000. This information was derived from a survey completed by all students upon graduation. The customer characteristics included in the model were also based on input from MBA program administrators and help explain students' background and career orientation. For estimation of the PS approach, we define customer-types by the first job the student took after graduation: (1) Investment Banker, (2) Corporate Finance, (3) Technology Manager, (4) General Manager, (5) Brand Manager, (6) Consultant, and (7) Other. An additional customer characteristic included in the model indicates whether the student has a technical bachelor's degree. A summary of the customer characteristics is presented in Table 1.

TABLE 1
SAMPLE CHARACTERISTICS

Products (Courses)		**Customers (Students)**	
Product-type sample shares		Customer-type sample shares	
Accounting	9%	Investment banker	15%
Finance	28%	Corporate finance	9%
Management of information systems	22%	IT manager	10%
Management	32%	General manager	12%
Marketing	9%	Brand manager	18%
		Consultant	19%
		Other	17%
Product attribute:		Customer characteristic:	
Mean course evaluation score		Students with technical	
(min=1 / max=5)	4.12	undergraduate degrees	34%

Model Specification

We consider three alternative model specifications for the PS approach. In the first specification, Model 1, we include only customer-type-specific constants. To further our understanding of a product's strength of attraction for different customer-types, the second model specification includes an additional customer characteristic. Because certain courses (such as quantitative courses) may have higher attraction for students with technical backgrounds, the second specification, Model 2, adds to Model 1 an additional dummy variable that indicates whether a particular student has a technical undergraduate degree. Finally, we consider the impact of a particular product attribute on the strength with which a product attracts different customer-types. Because courses that receive higher evaluations may attract different students, Model 3 adds to Model 2 a variable indicating the average evaluation score for each course[1]. Each of the model specifications for the PS approach is presented in Appendix E. We apply each of the three model specifications sequentially to evaluate the information provided by the additional predictors. Within each model specification, we systematically increase the number of product segments in the model and monitor the change in the log-likelihood and Consistent Akaike Information Criterion (CAIC). Across model specifications and product segment levels, we select the model with the lowest CAIC while maximizing the log-likelihood as the model with the best fit.

[1] Because some courses are offered more than once during our period of observation, we are not able to link students who take the course to a particular offering of a given course. Hence, we represent the evaluation of the course by the average evaluation the course received during the period of observation.

We also consider three alternative model specifications for the CS approach. In the first specification, Model 4, we include only product-type-specific constants. To further our understanding of a customer's preference for different product-types, the second CS model specification includes an additional product attribute. Because certain students may prefer courses with higher course evaluations, the second specification, Model 5, adds to Model 4 the variable that indicates the average course evaluation score for the particular course. Finally, we consider the impact of a particular customer characteristic on a customer's preference for different product-types. Because students who have technical backgrounds might prefer different courses, Model 6 adds to Model 5 the dummy variable indicating whether a student has a technical undergraduate degree. Each of the model specifications for the CS approach is presented in Appendix E. As with the PS approach, we sequentially apply each of the three specifications of the CS approach, systematically increase the number of customer segments within each model specification, and, across specifications, select the model with the lowest CAIC as the model with the best fit.

Product Segmentation Approach Results

Directly identified product segments. As indicated in Table 2, the 4-product segment solution of Model 1 has the lowest CAIC compared to other solutions for Models 1, 2, and 3. Hence we find that the explanatory power provided by the customer- and product-specific predictor variables in Models 2 and 3 was not great enough to overcome the cost of including those additional parameters. From Table 3, which summarizes the results of assigning courses to product segments using posterior probabilities, the 4-product segment solution of Model 1 indicates that Product Segment 1, "Quantitative" courses, is made up of some Finance, Accounting, and MIS courses; Product Segment 2, "Technical" courses, is exclusively made up of MIS courses; Product Segment 3, "Analytical" courses, is made up of Marketing courses with some Management and MIS courses; and Product Segment 4, "General Appeal" courses, is made up of Management courses with some Finance and Accounting courses. Thus, contrary to what one might expect, the identified segments indicate that all courses offered by a particular department do not attract the same types of students. For example, the MIS courses assigned to Product Segment 1 attract a different mix of students compared to the MIS courses assigned to Product Segment 2 or 3.

Customer-types attracted to each product segment. Table 3 also reports estimated attraction weights for the PS approach. Because the significance levels of the coefficients depend on the customer-type selected as the baseline, we cannot directly infer the mix of customers attracted to each product segment from those coefficients. Instead, we refer to the probabilities, prob($\Pi|C_0$) that are used to compare the relative effectiveness of product recommendations for the PS approach. We present those probabilities in the bottom section of Table 3 and highlight all instances in which the calculated recommendation probability exceeds the recommendation rule, (1/# product segments) = .25. That is, we can say that courses in product segment Π are more likely to be chosen by a customer of type C_0 than one would expect if that customer-type were choosing randomly among the four product segments. Combining this information with knowledge of the mix of courses in each product segment, the top half of Table 4 shows that, as we might expect, the "Quantitative" courses (Finance, Accounting, and

MIS) in Product Segment 1 attract students who pursue careers as Investment Bankers, Corporate Financiers, and Other careers; “Technical” courses (MIS) in Product Segment 2 attract students who pursue IT Management careers; and “Analytical” courses (Marketing, Management, and MIS) in Product Segment 3 attract students who pursue careers in Brand Management. However, the results also reveal findings that one may not have expected. In particular, we find that the “General Appeal” courses in Product Segment 4, which includes not only Management courses, but also Finance, and Accounting courses, attract all student-types.

As discussed earlier, the PS approach uses latent class analysis to directly identify product segments, but the approach can also be used to infer customer segments. To infer customer segments, we again refer to the probabilities used to compare the relative performance of the PS and CS approaches. In this case, we consider the probabilities, prob($C|P_0$) used to compare the relative effectiveness of customer target recommendations from the PS approach. We present those probabilities in Table 5 and highlight all instances in which the calculated recommendation probability exceeds the recommendation rule, (1/# customer-types) = .14. In these cases we say that courses of type P_0 attract customers of type C more strongly than one would expect if courses of type P_0 attracted all customer-types with equal strength.

Indirectly inferred customer segments. By observing the customer-types targeted by each of the product-types in Table 5, we can identify patterns in the recommendations across customer-types. We indirectly infer customer segments by grouping together customer-types for which we observe the same pattern of recommendations. As such, from the results of the PS approach we indirectly infer four customer segments: Investment Bankers and Corporate

TABLE 2
PRODUCT SEGMENTATION APPROACH: SELECTION OF BEST-FITTING MODEL

	Model 1 Customer-type-specific constants only					**Model 2 Customer-type-specific constants and customer characteristic**					**Model 3 Customer-type-specific constants, customer characteristic and product attribute**			
Model specification[a]														
No. segments	1	2	3	**4**	5	1	2	3	4	5	1	2	3	4
No. parameters	6	13	20	**27**	34	7	15	23	31	39	13	27	41	55
Predictor variables														
Customer-type-specific constants	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Technical undergrad degree						✓	✓	✓	✓	✓	✓	✓	✓	✓
Course evaluation score											✓	✓	✓	✓
Fit statistics														
Log-likelihood	-5296	-5133	-5072	**-5038**	-5035	-5295	-5126	-5070	-5036	-5034	-5283	-5109	-5057	-5025
CAIC	10646	10383	10323	**10317**	10382	10655	10387	10345	10350	10406	10682	10459	10481	10542

[a] Values in bold/underlined indicate the selected model specification

TABLE 3
PRODUCT SEGMENTATION APPROACH: PRODUCTS ASSIGNED TO SEGMENTS, CUSTOMER-TYPES ATTRACTED BY SEGMENTS, AND PRODUCT RECOMMENDATIONS

	Model 1: 4-Product Segment Solution			
Product-types	**Product Segment 1 "Quantitative" % of segment**	**Product Segment 2 "Technical" % of segment**	**Product Segment 3 "Analytical" % of segment**	**Product Segment 4 "General Appeal" % of segment**
Accounting courses	.20	.00	.00	.08
Finance courses	.70	.00	.00	.17
MIS courses	.10	1.00	.20	.00
Management courses	.00	.00	.20	.75
Marketing courses	.00	.00	.60	.00
Customer-types	**Estimated attraction weight[a]**	**Estimated attraction weight**	**Estimated attraction weight**	**Estimated attraction weight**
Investment banker	**.89**	**-1.90**	**-1.44**	-.14
Corporate finance	.21	**-1.08**	**-1.33**	**-.57**
IT manager	**-1.25**	**.68**	**-.90**	**-.65**
General manager	**-.63**	**-.52**	.03	-.16
Brand manager	**-1.00**	-.32	**.77**	.06
Consultant	.06	**.53**	.22	**.19**
Other	.00	.00	.00	.00
Product segment size	.30	.16	.16	.38
Customer-types	**Probability of recommending segment[b,c]**	**Probability of recommending segment**	**Probability of recommending segment**	**Probability of recommending segment**
Investment banker	**.63**	.02	.03	**.31**
Corporate finance	**.52**	.08	.06	**.34**
IT manager	.12	**.46**	.10	**.32**
General manager	.20	.12	.22	**.45**
Brand manager	.11	.11	**.35**	**.43**
Consultant	.24	.21	.16	**.39**
Other	**.29**	.15	.16	**.40**

[a] Estimated attraction weights in bold are significant at $p < .05$
[b] Recommendation probability = prob(Π | C_0)
[c] Recommendation probabilities in bold exceed recommendation rule (1 / # product segments) = .25

Financiers (who are targeted by Accounting and Finance courses), Consultants and Others (who are targeted by all course types),Brand Managers and General Managers (who are targeted by Management and Marketing courses), and IT Managers (who are targeted by MIS courses). We summarize the customer segments inferred from the PS approach in the bottom half of Table 4.

TABLE 4
PRODUCT SEGMENTATION APPROACH: SUMMARY OF PRODUCT SEGMENTS AND CUSTOMER SEGMENTS

Directly Identified Product Segments	**Product Segment 1 "Quantitative"**	**Product Segment 2 "Technical"**	**Product Segment 3 "Analytical"**	**Product Segment 4 "General Appeal"**
Product-types in segment	Accounting Finance MIS	MIS	Marketing MIS Management	Management Finance Accounting
Customer-types to which segment is recommended	Investment Bankers Corporate Finance Other	IT Managers	Brand Managers	All student-types
Indirectly Inferred Customer Segments	**Customer Segment 1 "Financiers"**	**Customer Segment 2 "Consultants"**	**Customer Segment 3 "General Mgrs"**	**Customer Segment 4 "Tech Mgrs"**
Customer-types in segment	Investment Bankers Corporate Finance	Consultants Other	Brand Managers General Managers	IT Managers
Product-types targeted to segment	Accounting Finance	All course-types	Management Marketing	MIS

TABLE 5
PRODUCT SEGMENTATION APPROACH: PROBABILITY OF RECOMMENDING CUSTOMERS FOR PRODUCT-TYPES TO TARGET

	Probability of Recommending Customer-types[a,b]						
Product-types	**Investment Banker**	**Corporate Finance**	**Other**	**Consultant**	**Brand Manager**	**General Manager**	**IT Manager**
Accounting	**.28**	**.15**	**.15**	**.17**	.09	.10	.06
Finance	**.30**	**.16**	**.15**	**.16**	.08	.09	.05
MIS	.07	.07	**.15**	**.24**	.13	.10	**.23**
Management	.13	.09	**.16**	**.20**	**.19**	**.14**	.08
Marketing	.04	.04	**.16**	**.20**	**.34**	**.16**	.06

[a] Recommendation probability = prob(C | P_0)
[b] Probabilities in bold exceed recommendation rule (1 / # customer-types) = .14

Customer Segmentation Approach Results

Directly identified product segments. As indicated in Table 6, the 2-customer segment solution of Model 5, which includes average course evaluation as a predictor, has the lowest CAIC relative to all other solutions for Models 4, 5, and 6. Hence we find that including course evaluation scores as a predictor helps explain students' preference for courses. From Table 7, which summarizes the results of assigning students to customer segments using posterior probabilities, the 2-customer segment solution of Model 5 indicates that Customer Segment 1 is primarily made up of "Quantitative" students, (Investment Bankers, Corporate Financiers, Consultants, and Others), while Customer Segment 2 is primarily made up of "Analytical" students (Brand Managers, General Managers, IT Managers, Consultants, and Others). Thus, contrary to what one might expect, the identified segments indicate that all students pursuing a particular career do not prefer the same types of courses. For example, the Consultants assigned to Customer Segment 1 prefer a different mix of courses compared to the Consultants assigned to Customer Segment 2.

TABLE 6
CUSTOMER SEGMENTATION APPROACH: SELECTION OF BEST-FITTING MODEL

	Model 4 Product-type-specific constants only				**Model 5 Product-type-specific constants and product attribute**				**Model 6 Product-type-specific constants, product attribute and customer characteristic**			
Model specification[a]												
No. segments	1	2	3	4	1	**<u>2</u>**	3	4	1	2	3	4
No. parameters	4	9	14	19	5	**<u>11</u>**	17	23	9	19	29	39
Predictor variables												
Product-type-specific constants	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Course evaluation score					✓	✓	✓	✓	✓	✓	✓	✓
Technical undergrad degree									✓	✓	✓	✓
Fit statistics												
Log-likelihood	-4144	-4025	-4001	-3998	-3695	**<u>-3634</u>**	-3627	-3624	-3694	-3630	-3624	-3621
CAIC	8323	8130	8127	8166	7435	**<u>7365</u>**	7405	7453	7469	7430	7506	7590

[a] Values in bold/underlined indicate the selected model specification

TABLE 7
CUSTOMER SEGMENTATION APPROACH: CUSTOMERS ASSIGNED TO SEGMENTS, PRODUCT-TYPES PREFERRED BY SEGMENTS, AND CUSTOMER RECOMMENDATIONS

	Model 5: 2-Customer Segment Solution	
Customer-types	**Customer Segment 1 "Quantitative" % of segment**	**Customer Segment 2 "Analytical" % of segment**
Investment bankers	.28	.02
Corporate finance	.16	.02
IT managers	.03	.18
General managers	.10	.14
Brand managers	.11	.24
Consultants	.16	.22
Other	.16	.18
Product-types	**Estimated preference weight[a]**	**Estimated preference weight**
Accounting	.00	.00
Finance	**1.00**	-.08
MIS	**-.60**	**.65**
Management	-.05	**.51**
Marketing	-.01	**.33**
Course evaluation	**4.62**	**3.48**
Customer segment size	.49	.51
Product-types	**Probability of recommending segment[b, c]**	**Probability of recommending segment**
Accounting	**.54**	.46
Finance	**.76**	.23
MIS	**.52**	.47
Management	.18	**.82**
Marketing	.12	**.87**

[a] Estimated preference weights in bold are significant at $p < .05$
[b] Recommendation probability = prob(χ | P_0)
[c] Recommendation probabilities in bold exceed recommendation rule (1 / # customer segments) = .50

Product-types preferred by each customer segment. Table 7 also reports estimated preference weights for the CS model. Because the significance levels of the product-type-specific coefficients depend upon the product-type selected as the baseline, we cannot directly infer the types of courses preferred by each customer segment from those coefficients. Instead, we refer to the probabilities, prob(χ|P_0), which are used to compare the relative effectiveness of customer target recommendations from the CS approach. We present those probabilities in the bottom section of Table 7 and highlight all instances in which the calculated recommendation probability exceeds the recommendation rule, (1/# customer segments) = .50. That is, we can say that courses of type P_0 attract customers from segment χ more strongly than one would expect if courses of type P_0 attracted all customer segments with equal strength. Combining this information with knowledge of the mix of students in each customer segment in the top half of Table 8, shows that, as we might expect, "Quantitative" students in Customer Segment 1 prefer Accounting, Finance, and MIS courses, while "Analytical" students in Customer Segment 2 prefer Marketing and Management courses.

Indirectly inferred product segments. As discussed earlier, the CS approach uses latent class analysis to directly identify customer segments, but this approach can also be used to infer product segments. To infer product segments, we again refer to the probabilities used to compare the relative performance of the PS and CS approaches. This time we consider the probabilities, prob(P|C_0), used to compare the relative effectiveness of product recommendations from the CS approach. We present those probabilities in Table 9 and highlight all instances in which the calculated recommendation probability exceeds the recommendation rule, (1/#

product-types) = .20. That is, we say that products of type P more strongly attract customers of type C_0 than one would expect if products of type P attracted all customer-types equally.

By observing the product-types recommended to each of the customer-types in Table 9, we can identify patterns in the recommendations across product-types. We indirectly infer product segments by grouping together product-types for which we observe the same pattern of recommendations. As such, with the results of the CS approach we indirectly infer three product segments: Management courses (which are recommended to all customer-types), Finance courses (which are recommended to all customer-types except IT Managers), and MIS courses (which are recommended to IT Managers and Brand Managers). We summarize the product segments inferred from the CS approach in the bottom half of Table 8.

TABLE 8
CUSTOMER SEGMENTATION APPROACH: SUMMARY OF CUSTOMER SEGMENTS AND PRODUCT SEGMENTS

Directly Identified Customer Segments	**Customer Segment 1 "Quantitative"**	**Customer Segment 2 "Analytical"**
Customer-types in segment	Investment bankers Corporate financiers General managers Consultants Others	Brand managers IT managers General managers Consultants Others
Product-types targeted to segment	Accounting Finance MIS	Marketing Management

Indirectly Inferred Product Segments	**Product Segment 1 "Finance"**	**Product Segment 2 "MIS"**	**Product Segment 3 "Management"**
Product-types in segment	Finance	MIS	Management
Customer-types to which segment is recommended	All student-types except IT managers	Brand managers IT managers	All student-types

TABLE 9
CUSTOMER SEGMENTATION APPROACH: PROBABILITY OF RECOMMENDING PRODUCTS TO CUSTOMER-TYPES

	Probability of Recommending Product-types[a, b]				
Customer-types	**Accounting**	**Finance**	**MIS**	**Management**	**Marketing**
Investment banker	.18	**.43**	.06	**.27**	.06
Corporate finance	.17	**.42**	.07	**.27**	.06
Other	.16	**.28**	.17	**.26**	.13
Consultant	.16	**.27**	.18	**.26**	.14
Brand manager	.15	**.23**	**.21**	**.25**	.15
General manager	.16	**.26**	.18	**.26**	.14
IT Manager	.15	.18	**.25**	**.25**	.18

[a] Recommendation probability = prob(P | C_0)
[b] Probabilities in bold exceed recommendation rule (1 / # product-types) = .20

Comparing Product Segmentation Approach and Customer Segmentation Approach Results

Managerial comparison of approaches' segmentations. The differences between the directly estimated product segments from the PS approach and the inferred product segments from the CS approach are illustrated by comparing the top half of Table 4 and the bottom half of Table 8. This comparison highlights the relationship between the product segmentation schemes derived from the two approaches. The inferred product segment "Management" courses is roughly analogous to directly identified Product Segment 4 "General Appeal" courses. The inferred "Finance" courses product segment is roughly analogous to directly identified Product Segment 1 "Quantitative" courses. The inferred product segments "MIS" courses is roughly analogous to directly identified Product Segments 2 and 3, "Technical" courses and "Analytical" courses. While we can see a rough equivalence, it is very likely that we will get better product recommendations from the product segments directly constructed by the PS approach since the composition of those product segments is not constrained to take “all or none” of the courses offered by a particular department as are the inferred product segments from the CS approach.

Similarly, the differences between the directly estimated customer segments from the CS approach and the inferred customer segments from the PS approach are illustrated by comparing the top half of Table 8 and the bottom half of Table 4. This comparison highlights the relationship between the customer segmentation schemes derived from the two approaches. The inferred customer segment “Financiers” is roughly analogous to directly identified Customer Segment 1 “Quantitative” students. The inferred customer segments “General Managers” and “Technology Managers” are roughly analogous to directly identified Customer Segment 2 “Analytical” students. The inferred customer segment “Consultants and Others” is split between Customer Segment 1 and Customer Segment 2. Thus, although we see a rough equivalence, it is again very likely that we will get better customer target recommendations from the customer segments directly estimated by the CS approach since the composition of those customer segments was not constrained to take “all or none” of a particular customer-type as are the inferred customer segments from the PS approach.

Empirical comparison of approaches’ recommendations. To assess the relative performance of the PS and CS approaches, we first apply both approaches to recommend a set of courses for a withheld student to take and compare the approaches’ recommendations with the set of courses the withheld student actually took. The hit rate used in this approach comparison is the proportion of courses the withheld student actually took that the approach recommended. In a second test, we apply both approaches to recommend a set of students for a withheld course to target and compare the approaches’ recommendations with the set of students who actually took the withheld course. The hit rate for this test reports the proportion of students who actually took the withheld course that the approach recommended as targets[2].

In the first test, the PS approach provided statistically significantly better recommendations of courses for a withheld student to take (PS hit rate = .63, CS hit rate = .45, $p < .01$ based on a paired sample t-test).   In the second test, the CS approach provided better recommendations of students for a withheld course to target (CS hit rate = .46, PS hit rate = .41, difference not significant based on paired sample t-test). Note that the CS approach's recommendations of students for a withheld course to target were statistically significantly better than recommendations from the PS approach using the alternate recommendation rules.

CHAPTER 5: DISCUSSION

We began this research with two primary objectives. Our first objective was to develop a product segmentation approach that could be applied by retailers with large multi-category product offerings to identify latent groups of products. We refer to these groupings as product segments such that the products within a segment attract the same types of customers, while products in different segments attract different types of customers. Our second objective was to examine the relationship between the product segmentation approach and a parallel customer segmentation approach by comparing the segments identified by each approach and the

[2] We note that both estimated models were stable through the n-fold bootstrapping procedures. Across each of the 326 withheld students in the first test, the structure of neither the best fitting PS model nor the best fitting CS model changed. Similarly there were no changes in the structure of the best fitting models across any of the 32 withheld courses in the second test. In Appendix F, we present statistics that speak to the stability of the estimated models.

effectiveness of each approach at addressing two managerial questions: Which products should be recommended to a customer? Which customers should a product target?

We addressed our first objective by applying the proposed product segmentation approach to identify latent product segments among courses offered by a business school. To begin estimation of the product segmentation approach, business school administrators defined seven managerially relevant customer-types based on the careers students pursue after graduating (Investment Bankers, Corporate Finance, IT Manager, Brand Manager, Consultant, and Other). Given the defined customer-types, the product segmentation approach directly identified four product segments: “Quantitative” courses, “Technical” courses, “Analytical” courses, and “General Appeal” courses, where each product segment was made up of courses from several different departments. Thus, the product segmentation approach provides a methodology for retailers with large, multi-category product and service offerings to directly identify customer-centric, cross-category product groupings.

We addressed the second objective by comparing results identified by the proposed product segmentation approach with results from a parallel customer segmentation approach. To begin estimation of the customer segmentation approach, business school administrators defined five managerially relevant product-types based on the business school’s departments (Accounting, Finance, MIS, Management, and Marketing). Given the defined product-types, the customer segmentation approach directly identified two customer segments: “Quantitative” students and “Analytical” students, where each customer segment was made up of students who pursued different types of careers.

Comparing results from the product segmentation approach and the customer segmentation approach in our illustrative application, we see the implication of representing students by only seven customer-types – the aggregation constraint that makes estimation of the product segmentation approach tractable – and the impact of representing courses by five product-types – the aggregation constraint that makes estimation of the customer segmentation approach tractable. The four product segments directly identified by the product segmentation approach are not identical to the three product segments indirectly inferred from the results of the customer segmentation approach. Similarly, the two customer segments directly identified by the customer segmentation approach are not identical to the four customer segments indirectly inferred from the results of the product segmentation approach. As such, while the proposed product segmentation approach parallels the widely applied latent class customer segmentation approach, our application illustrates the fact that the two approaches are not exactly "reverse sides of the same analysis" as suggested by Grover and Srinivasan (1987). Rather, the aggregation constraint imposed on customers in the product segmentation approach and imposed on products in the customer segmentation approach influence the product segments and customer segments that each approach identifies. Thus, we contribute to marketers' understanding of the relationship between customer segmentation and product segmentation by illustrating the implications of the aggregation constraint in the underlying models.

Further, the aggregation constraints degrade model performance. Specifically, we see that the aggregation constraint imposed on customers in the product segmentation approach degrades that approach's recommendations of students for a course to target. Similarly, we see that the aggregation constraint imposed on products in the customer segmentation approach degrades that approach's recommendations of courses for a student to take. Thus, the decision of whether to apply the product segmentation approach or the customer segmentation approach should be based on the particular managerial objectives involved in aligning a retailer's product offerings with its target customers.

Having demonstrated the usefulness of the proposed product segmentation approach in a service provider context, the potential for further application is clear. Retailers with large multi-category product or service offerings that have customer management systems can use the power of latent class analysis coupled with the simplicity and flexibility of the multinomial logit-like structure of the proposed product segmentation approach to extract insights from their data and to guide managerial action. For example, Wal-Mart, having already identified "Hispanics, " "African-Americans," "Suburbanites," "Rural Residents," "Affluent," and "Empty-nesters" as its target customer-types, could use this methodology to design store layouts that organize products into groups that attract a particular customer-type. Direct retailers such as Amazon and Dell could apply the methodology to dynamic website design whereby, when a customer of a particular customer-type logs on to the website, the web page features the set of products most likely to attract that customer-type. Retailers such as Best Buy could apply the approach to develop targeted direct mail campaigns that offer promotions on products from the set of items that is most likely to attract a particular customer-type. The product segmentation approach could help retailers cross-merchandise by identifying which products to display together or for salespeople to recommend to particular customer-types.

Limitations and Future Research

While the proposed product segmentation approach presents a parsimonious approach for identifying product segments from large numbers of products in multiple categories, it also has limitations that invite further model development opportunities. First, analogous to the IIA assumption in logit choice models, the attraction model that underlies the product segmentation approach implicitly assumes that adding a new customer-type to a product's customer mix will not change the relative strength of attraction that the product has for existing customer-types. It is easy to imagine a scenario in which such an assumption will not hold. Consider the case in which a product becomes attractive to a new customer-type that is an opinion leader (e.g. media personalities). The arrival of that new customer-type in a product's customer mix might increase the product's relative strength of attraction for existing, impressionable customer-types and might decrease the product's relative strength of attraction for existing customer-types who tend to avoid fads. Models that relax the IIA assumption could be applied to remedy this limitation.

Second, analogous to the assumption in logit choice models that repeated choices over time are independent, the attraction model that underlies the product segmentation approach assumes that a product's strength of attraction for a particular customer-type is independent of the other customer-types attracted. It is also possible to imagine a scenario in which this will not hold. Customer-types that are closely related to each other may influence each other's attraction probability in a way not captured by the attraction model. Again, this limitation could be addressed by incorporating model developments designed to relax this assumption.

Key results in this research rest on the impact of imposing the aggregation constraint on customers in the proposed product segmentation approach and imposing the aggregation constraint on products in the customer segmentation approach. Since imposing the product-type aggregation constraint has become standard procedure in estimating choice models, the implications of the constraint has received little consideration. In this research, we demonstrate some of the implications of imposing the product-type aggregation constraint and highlight the need for marketers to give wider consideration to the impact of the a priori assumptions used to impose product-types. It would be valuable to compare this approach of identifying product segments and customer segments with approaches that do not impose an aggregation constraint on either customers or products, such as non-parametric methods.

Finally, the characteristics of the data in our illustrative application also entail limitations that open opportunities for future research. In our application, we had no record of customers' responses to marketing interventions. As such, we were unable to capture the impact of marketing interventions on the strength with which a product attracts different customer-types and the impact of changes in that attraction strength on product segmentation. It would be interesting to apply the proposed product segmentation approach in a context that incorporates marketing activities to assess their impact on product segmentation. In general, we invite application of the proposed approach in additional contexts that involve other large, multi-category collections of products, as well as observations of the interventions used to market those items.

ESSAY 2: A COMPARISON OF ATTRIBUTE-BASED CO-CLUSTERING, LATENT CLASS SEGMENTATION, AND COLLABORATIVE FILTERING RECOMMENDATION SYSTEMS

CHAPTER 6: INTRODUCTION

In both traditional and online formats, many retailers present customers with very large product or service offerings. Category specialist retailers such as Best Buy and Barnes and Noble typically offer more than 20,000 products in their brick-and-mortar stores (Levy and Weitz 2006), while the online channel allows for virtually unlimited shelf space. Although access to large product or service offerings may be desirable for customers, the large offerings also make it difficult for customers to process information about all of the choice alternatives (Haubl and Trifts 2000). The complexity of the choice set and uncertainty about the set of possible alternatives can delay or prevent customers' purchasing (West et al. 1999). In recent years, recommendation systems have become an increasingly important tool for customer-centric retailers with very large product or service offerings, who employ these "intelligent agents" in an effort to help customers search and choose from their offerings. Prominent examples of retailers who widely apply recommendation systems include Amazon.com, Barnes and Noble, Blockbuster, and Netflix. An entire industry has developed to provide retailers such as these with recommendation system solutions designed to help the retailers better serve their customers.

In retailing contexts, recommendation systems represent methodologies for making product recommendations to a customer based on an analysis of the characteristics and prior behavior of that customer, as well as information on other customers. The roles of recommendation systems include helping customers construct preferences, find and organize relevant information, and evaluate and purchase attractive alternatives (West et al. 1999). In such roles, product recommendation systems can be implemented in a number of activities including designing customized websites and developing targeted promotions. Extant research indicates that, in online settings, recommendation systems can help consumers make better quality decisions while expending less effort (Haubl and Trifts 2000). As such, recommendation systems have the potential to increase customer repurchasing and satisfaction and improve customer retention and loyalty. To achieve these objectives, it is important for retailers to employ recommendation systems that provide high quality recommendations.

While the evaluation and enhancement of recommendation system performance has been widely studied in computer science and machine learning over the past decade, most extant research has applied collaborative filtering approaches such as nearest neighbor algorithms. Recent developments in attribute-based co-clustering algorithms represent a promising new approach that has not been empirically evaluated in terms of its performance in recommendation system settings. Co-clustering is a technique that simultaneously clusters items such as customer's product choices along two dimensions such as customers and products. The method has been applied in several domains including text clustering (Dhillon, Mallela and Modha 2003) and microarray data analysis (Cheng and Church 2000; Cho, Dhillon, Guan and Sra 2004). Recently, co-clustering algorithms have been shown to perform better than traditional collaborative filtering techniques in recommendation system contexts (George and Merugu 2005). However, the co-clustering algorithms previously proposed make recommendations based only on observed ratings or choices. In this research, we extend prior co-clustering algorithms to present a new, augmented approach that incorporates customer characteristics and product attributes, in addition to the observed ratings or choices, in the identification and recommendation of co-clusters. As such, the proposed attribute-based co-clustering approach represents a clustering method that yields insight into the customer characteristics and product attributes that drive the derived recommendations. We evaluate the performance of the proposed attribute-based co-clustering approach in a recommendation system application.

To evaluate the performance of attribute-based co-clustering as a recommendation system, we compare the proposed approach to two other recommendation system techniques. We first compare attribute-based co-clustering with a related model-based approach. Because recommendation systems are a relatively new concept in marketing (e.g., Ansari, Essegaier and Kohli 2000; Ying, Feinberg and Wedel 2006), little is known about how widely applied models of customer purchasing behavior such as latent class segmentation can be applied as recommendation systems. Murthi and Sarkar (2003) suggest that significant opportunities exist to evaluate the effectiveness of traditional approaches for modeling customer preferences as personalization techniques, while Ansari, Essegaier and Kohli (2000) specifically suggest that latent class models are worthy of investigation as recommendation systems. Godfrey, McAlister and Saar-Tsechansky (2007) find that latent class product segmentation and customer segmentation models that identify segments of products and customers, respectively, based on information on managerially relevant customer characteristics and product attributes can be applied as recommendation systems. However, no prior research has evaluated the relative performance latent class segmentation approaches to other types of recommendation systems. In this research, we evaluate and compare the performance of latent class segmentation to that of the proposed attribute-based co-clustering approach.

To further evaluate the performance of the proposed attribute-based co-clustering approach, it is important to also evaluate the technique against a widely applied and well-established recommendation system. The most commonly applied recommendation system approach is collaborative filtering and one of the most popular collaborative filtering algorithms is the nearest neighbor approach. The nearest neighbor algorithm is often referred to as a "memory-based" approach because no concise model is induced from the data, and the entire database is used each time a prediction is generated. In contrast, attribute-based co-clustering and latent class segmentation are "model-based" approaches. While several researchers have evaluated the relative performance of memory-based and model-based recommendation systems (e.g., Ansari, Essagaier and Kohli 2000; Ariely, Lynch and Aparicio 2004; Breese, Heckerman and Kadie 1998; Canny 2002; Chien and George 1999; Mild and Natter 2002; Mild and Reutterer 2003) it is difficult to generalize about the relative performance of these different classes of recommendation systems across prior research because the data samples of the various comparisons differ, even when the same data set (e.g. EachMovie data) is used. Features of the data set such as the sample size, the number of choice alternatives, and the density of ratings or choices have been shown to impact the relative performance of different algorithms (Herlocker, Konstan, Terveen, and Riedl 2004). However, little research systematically examines the impact of these data features on the relative performance of different recommendation systems. Further, prior research rarely articulates the impact of the predictive power of the customer characteristics and product attributes on the quality of recommendations for approaches that incorporate this information compared to approaches that do not. In this research, we consider the impact of each of these factors on the performance of the three types of recommendation systems.

This research contributes to marketers' understanding of recommendation systems in three respects. First, we present a new recommendation system methodology, attribute-based co-clustering. The proposed attribute-based co-clustering approach extends prior co-clustering algorithms to incorporate information on customer characteristics and product attributes in the recommendations. Second, we evaluate the relative performance of attribute-based co-clustering by comparing it to a related model-based approach, latent class segmentation, and a widely applied memory-based approach, nearest neighbor collaborative filtering. Finally, we investigate factors that impact the relative performance of the three types of recommendation systems. Specifically, we examine the impact of the quality of information on customer characteristics and product attributes and the impact of inherent features of the data set. We empirically evaluate the impact of these factors on the relative performance of the three approaches by comparing the quality of recommendations made by the three different approaches. We compare the quality of recommendations in two contexts relevant to retailers: recommending a set of products for a customer to purchase and recommending a set of customers for a product to target.

In the following sections, we review relevant research on recommendation systems to assess extant understanding on different types of recommendation systems and the factors that impact their performance. We present the proposed attribute-based co-clustering approach and compare that methodology with a related latent class product and customer segmentation approach and the widely applied nearest neighbor collaborative filtering approach. In an education service setting, we apply the three approaches in two recommendation tests: (1) recommend a set of products for a customer to purchase, and (2) recommend a set of customers for a product to target. We then evaluate the relative quality of the recommendations made by each approach in the two tests and discuss the contexts in which each approach performs best. We conclude by discussing implications of the results and suggesting directions for future research.

CHAPTER 7: LITERATURE REVIEW

Since Goldberg, Nichols, Oki and Terry (1992) first introduced collaborative filtering as a recommendation system technique, a significant stream of research has developed in computer science and machine learning that aims to refine and enhance the performance of collaborative filtering approaches (e.g., Herlocker, Konstan, Borchers and Riedl 1999; Konstan et al. 1997). Recently, as researchers and practitioners have recognized the relevance of recommendation systems to a variety of domains, a stream of research has developed that compares the relative performance of some well-established, memory-based approaches with model-based approaches. Several comparisons between memory-based and model-based approaches evaluate the relative quality of collaborative filtering recommendations against recommendations from regression-based models. Ariely, Lynch and Aparicio (2004) find that a logistic regression-based model performs better than k-mean and nearest neighbor collaborative filtering at predicting a customer's purchases using simulated data that includes scores on three positively valenced attributes for each of the 20 products. Other research finds that correlation-based collaborative filtering performs better than linear and logistic regression-based models using data describing ordinal ratings of movies (Mild and Natter 2002) and outperforms a binary logit-based model at predicting customer's market baskets in grocery transaction data (Mild and Reutterer 2003). In these comparisons, the model-based approaches are limited in that the regression models do not allow for customer heterogeneity.

While the relative performance of regression model-based recommendation systems that do not allow for heterogeneity tends to vary, prior research indicates that model-based approaches that incorporate heterogeneity tend to perform better than memory-based, collaborative filtering recommendation systems (e.g., Ansari, Essagaier and Kohli 2000; Breese, Heckerman and Kadie 1998; Chien and George 1999; Ying, Feinberg and Wedel 2006). Prior comparisons that incorporate customer heterogeneity in model-based approaches generally do so using Bayesian frameworks to analyze explicit, ordinal ratings data. The model-based approaches have included Bayesian mixture models (Chien and George 1999), Bayesian networks (Breese, Heckerman and Kadie 1998), and hierarchical Bayes models (Ansari, Essagaier and Kohli 2000; Ying, Feinberg and Wedel 2006). Ying, Feinberg and Wedel (2006) note that hierarchical Bayes models have been shown to offer roughly equal performance to finite mixture models in modeling customer heterogeneity in various empirical settings. However, little prior research has evaluated the relative performance of finite mixture models, or latent class segmentation models, as model-based recommendation systems.

One exception is Hoffman and Puzicha (1999), who apply a latent class approach to incorporate customer heterogeneity in analyzing explicit, ordinal movie ratings data. They convert the ratings data into binary chose / did not choose data and apply a maximum likelihood estimated latent class model. A second exception is Godfrey, McAlister and Saar-Tsechansky (2007), who find that latent class product segmentation and customer segmentation models that identify segments of products and customers, respectively, based on information on managerially relevant customer characteristics and product attributes can be applied as recommendation systems. These authors find that the MLE latent class models perform well as recommendation systems, however they do not compare the performance of this approach with other types of recommendation systems such as memory-based collaborative filtering algorithms. In this research, we compare the performance of latent class product segmentation and customer segmentation model-based recommendation system approaches with memory-based collaborative filtering. As such, one contribution of this research is to evaluate the performance of a latent class segmentation model-based approach of capturing heterogeneous customer purchasing behavior as a recommendation system, relative to other types of approaches.

Latent class segmentation recommendation systems cluster a sample along one dimension such as customers and make recommendations based on the revealed preferences of each segment. Co-clustering approaches, which cluster data along two dimensions such as customers and product simultaneously, present another recommendation system approach. Hoffman and Puzicha (1999) present a latent class-based co-clustering recommendation system approach. Although the authors find that the latent class-based co-clustering does not perform as well as a traditional, one-dimensional latent class segmentation approach, they do not compare the co-clustering method with other types of recommendation systems such as collaborative filtering. Recently, George and Merugu (2005) show that partitional co-clustering algorithms, which cluster the data matrix in to m customer clusters and n product clusters, perform better than traditional collaborative filtering techniques in recommendation system contexts. However the co-clustering algorithm evaluated makes recommendations based only on observed ratings or choices. In this research, we extend prior co-clustering algorithms to present a new approach that incorporates customer characteristics and product attributes, in addition to the observed ratings or choices, in the identification and recommendation of co-clusters. As such, the proposed attribute-based co-clustering approach represents a clustering method that yields insight into the customer characteristics and product attributes that drive the derived recommendations. An important contribution of this research is to evaluate the performance of the proposed attribute-based co-clustering in a recommendation system application.

Within the stream of research that has compared the performance of model-based and memory-based recommendation systems (e.g., Ansari, Essagaier and Kohli 2000; Ariely, Lynch and Aparicio 2004; Breese, Heckerman and Kadie 1998; Canny 2002; Chien and George 1999; Mild and Natter 2002; Mild and Reutterer 2003), the factors that impact relative recommendation quality across approaches are not well articulated. Of this research, almost all comparisons of model-based and memory-based recommendation systems use the EachMovie data set, which describes users' explicit ordinal ratings of movie titles. While this collection of research is applied to the same data set, the findings are equivocal in terms of the relative performance of model-based and memory-based approaches (see Table 10). Several findings indicate that model-based approaches out-perform memory-based approaches (e.g., Ansari, Essagaier and Kohli 2000; Canny 2002; Chien and George 1999; Ying, Feinberg and Wedel 2006). In contrast, other comparisons using the EachMovie data suggest that memory-based approaches outperform model-based approaches under certain conditions. Breese, Heckerman and Kadie (1998) find that correlation-based collaborative filtering and vector similarity-based collaborative filtering perform better than a Bayesian mixture model and a Bayesian network model when the approaches are given little information. Mild and Natter (2002) show that correlation-based collaborative filtering performs better than linear regression models when the number of customers is low as well as when the density of ratings is high.

The results across these studies in terms of the relative performance of model-based and memory-based approaches are equivocal in part because the sample features of the data vary in each study. In particular, Herlocker et al. (2004) note that three sample features can lead to different relative performance among algorithms. Specifically, Herlocker et al. (2004) suggest that three sample features that impact the relative performance of different recommendation systems include (1) the density of the ratings set overall, e.g., the average number of products chosen per person, (2) the density of ratings for the target customer, e.g., the number of observed choices for the target customer relative to the total number of choice alternatives, and (3) the size and distribution of the data set, e.g., sample size and number of choice alternatives. Further, some of the recommendation system approaches include customer characteristics and product attributes (Ansari, Essagaier and Kohli 2000; Ying, Feinberg and Wedel 2006), while other approaches do not. The impact of the inclusion and quality of the customer characteristics and product attributes on the relative performance of the approaches is generally not articulated. In this research, we systematically examine the impact of the quality of the customer characteristics and product attribute data as well as data sample features on the relative performance of the three types of recommendation systems. As such, a further contribution of this research is to highlight the impact of these factors on the quality of recommendations for different recommendation systems.

In summary, this research addresses gaps in prior literature that evaluates different classes of recommendation systems in the following ways. We present a new model-based recommendation system, attribute-based co-clustering, and evaluate its performance by empirically comparing the quality of its recommendations with other approaches. We examine the relative performance of attribute-based co-clustering by comparing recommendation quality with latent class segmentation, a related model-based approach which has previously been applied as a recommendation system but has not been evaluated relative to other approaches, and nearest neighbor collaborative filtering, a widely applied memory-based approach. In our comparison, we systematically examine the impact of the type of information and the sample features of the data on each approach. As such, we shed insight into the factors that impact the relative performance of three different types of recommendation systems, which has not comprehensively been articulated in the extant research.

TABLE 10
REVIEW OF PRIOR COMPARISONS OF MODEL-BASED AND MEMORY BASED RECOMMENDATION SYSTEMS

	Classes Compared		Data Type		Incorporates Customer Heterogeneity	Class with Best Recommendations
Authors	Model-Based	Memory-Based	Explicit (Ratings)	Implicit (Choices)		
Ansari, Essegaier and Kohli (2000)	X	X	X		X	Model-based
Ariely, Lynch and Aparicio (2004)	X	X		X		Model-based
Breese, Heckerman and Kadie (1998)	X	X	X	X	X	Model-based / Memory-based
Canny (2002)	X	X	X			Model-based
Chien and George (1999)	X	X	X		X	Model-based
George and Merugu (2005)	X	X	X			Model-based
Hoffman and Puzicha (1999)	X		X		X	Model-based
Mild and Natter (2002)	X	X	X			Memory-based
Mild and Reutterer (2003)	X	X		X		Memory-based
Ying, Feinberg and Wedel (2006)	X		X		X	Model-based
Current Research	X	X		X	X	

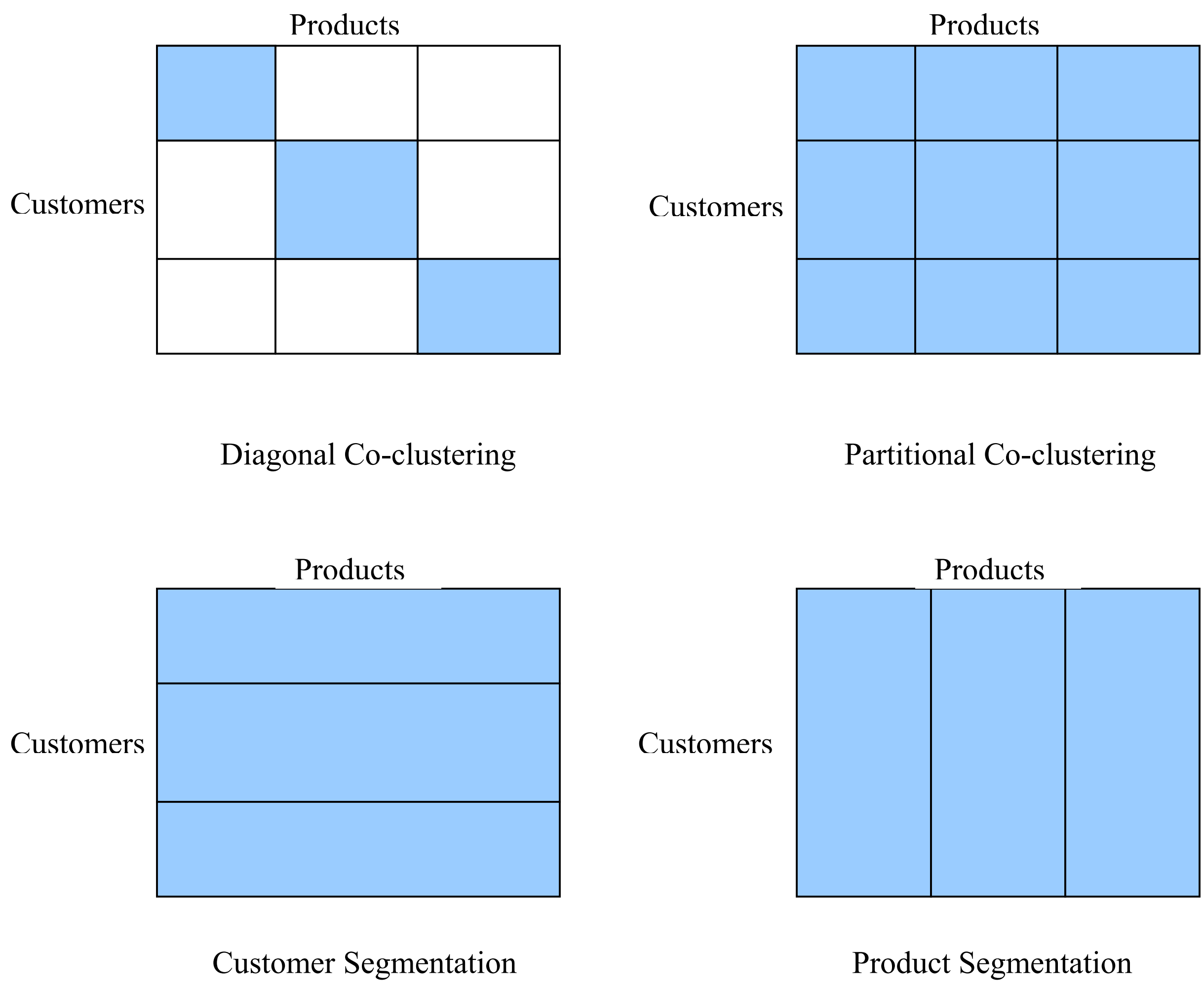
CHAPTER 8: METHODOLOGY

Attribute-based Co-clustering

Assume a data set where each row represents a customer and each column a product. Further assume a cell value in this matrix is "1" if a customer bought the product in the past and "0" otherwise. In traditional clustering, one either clusters customers into cohesive groups (based on their prior purchases) or, alternatively, clusters products into cohesive groups (based on the customers who purchased the products). Co-clustering, also known as bi-clustering, is a technique that clusters along two axes such as customers and products simultaneously. It exploits the duality between the axes and has been shown to give better results than one-sided clustering on several datasets. The approach has been successfully applied in domains such as text clustering (Dhillon, Mallela and Modha 2003) and microarray data analysis (Cheng and Church 2000; Cho et al. 2004). The earliest reference to co-clustering in the marketing literature is Wedel and Steenkamp (1991), who develop a generalized fuzzy clusterwise regression algorithm based on customers' product preferences. More recently, Hoffman and Puzicha (1999) develop a latent class-based co-clustering model. However the hard version of Wedel and Steenkamp's (1991) fuzzy clusters and Hoffman and Puzicha's (1999) latent co-clusters have a diagonal structure while the proposed attribute-based co-clustering approach is partitional and develops a full grid of k customer clusters and l product clusters. The distinction between the two types of co-clustering approaches is illustrated in Figure 1. In this figure, we graphically represent the results identified by Wedel and Steenkamp (1991) and Hoffman and Puzicha

(1999) in the diagram titled “Diagonal Co-Clustering,” where the matrix of customers and products is grouped into the diagonal clusters of a grid of k customer clusters and l product clusters. In contrast, we graphically represent the co-clusters identified by the proposed attribute-based approach in the diagram titled “Partitional Co-clustering,” where the matrix of customers and products can be clustered into full grid of co-clusters.

FIGURE 1
GRAPHICAL REPRESENTATION OF CLUSTERING FOR DIFFERENT RECOMMENDATION SYSTEM APPROACHES



While there are many co-clustering approaches, the proposed attribute-based co-clustering approach is based on Bregman co-clustering (Banerjee et al. 2005). Bregman co-clustering is a generalized framework for partitional co-clustering; it partitions the data matrix into a grid of k row and l column clusters. Bregman co-clustering can also be thought of as a matrix approximation technique that approximates the data matrix by a set of statistics preserved by each co-cluster. There are six possible schemes or bases that preserve different summary statistics of the data matrix, each of which produces a reconstructed matrix using the co-clustering solution. Attribute based co-clustering is an extension of basis 2, also known as Bregman block average co-clustering, which approximates all the values within a co-cluster by the mean of the values and hence preserves the co-cluster means.

Recently co-clustering has been applied in a recommendation system setting. In recommendation system contexts, the data is in the form of a matrix where the rows are customers, the columns are products, and the cell values are ratings or binary choices. The known ratings are used to simultaneously cluster customers and products and compute summary statistics for the co-clusters, which are then used to predict unknown ratings or choices. George and Merugu (2005) use an instance of the Bregman co-clustering algorithm on the EachMovie dataset to predict unknown user-movie ratings.

Bregman co-clustering uses only the customer-product matrix of ratings to make recommendations and does not incorporate potentially useful information about customer characteristics and product attributes. When such information is available, it may be possible to use this information to improve clustering. We propose an attribute based co-clustering algorithm, which extends a special case of the Bregman co-clustering algorithm (basis 2), to make use of additional information on customer characteristics and product attributes. Ordinary Bregman co-clustering has a collaborative filtering flavor since it uses only prior choices to cluster values and then assign a new customer or new product to the most likely co-cluster. In contrast, attribute-based co-clustering can make recommendations for new customers and products by comparing the similarity of new customers or products to existing ones based on attribute information in addition to prior choice information. The method of assigning a customer or a product to a co-cluster is described in the following section.

Identify attribute-based co-clusters. Letting m be the total number of customers and n be the total number of products, then Z is an m x n matrix of customers and products with cells representing the corresponding customer-product preference values. The preference values can be ratings or binary choices. Each customer u has characteristics $\mathbf{C_u}$, and each product v has attributes $\mathbf{P_v}$. We assume that the customer characteristics and product attributes are standardized to have the 0 mean and unit variance. Each customer-product preference in matrix Z corresponds to a vector $\mathbf{x_{uv}} = [z_{uv}, \mathbf{C_u}^T, \mathbf{P_v}^T]^T$ consisting of the actual matrix cell value z_{uv} (rating or binary choice) and the attributes for the corresponding customer and product. The distance between two customer-product choices (a-b, c-d) is now calculated as the distance between the corresponding vectors $\mathbf{x_{ab}}$, $\mathbf{x_{cd}}$. This distance is a weighted combination of the distances between the cell values and the customer and product attributes.

(1) $$d(\mathbf{x_{ab}}, \mathbf{x_{cd}}) = sqE\ (z_{ab}, z_{cd}) + w_1 sqE(\mathbf{C_a}, \mathbf{C_c}) + w_2 sqE(\mathbf{P_b}, \mathbf{P_d})$$

where the distance between the cell values and the attribute vectors is squared Euclidean distance. That is,

(2) $$d(\mathbf{x_{ab}},, \mathbf{x_{cd}}) = (z_{ab} - z_{cd})2 + w_1(\mathbf{C}_a - \mathbf{C_c})^T(\mathbf{C_a} - \mathbf{C_c}) + w_2(\mathbf{P_b} - \mathbf{P_d})^T(\mathbf{P_b} - \mathbf{P_d})$$

In Bregman block average co-clustering, the summary statistic preserved for each co-cluster is the co-cluster mean. In this extension of Bregman block average co-clustering, the aim is to simultaneously cluster the customers into k row clusters and the products into l column clusters such that the summary statistic preserved for each co-cluster is the mean vector of the preference value and the customer and product attributes. Let ρ be a mapping from the m customers to the k row clusters and γ be a mapping from the n products to the l column clusters. We want to find a co-clustering (ρ, γ), that minimizes the objective function

(3)
$$\sum_{g=1}^{k}\sum_{h=1}^{l}\sum_{u:\rho(u)=g}\sum_{v:\gamma(v)=h} d(x_{uv}, x_{uv}')$$

where $\mathbf{x_{uv}}$ is a vector consisting of the cell value z_{uv} and the corresponding customer attributes, $\mathbf{C_u}$, and product attributes, $\mathbf{P_v}$. $\mathbf{x_{uv}}'$ is the predicted vector, which is the mean vector for the co-cluster g-h, consisting of the mean cell value, μ_{gh}, the mean customer attributes $\mathbf{C\mu_{gh}}$, and the mean product attributes, $\mathbf{P\mu_{gh}}$ ($\mathbf{x_{uv}}' = [\mu_{gh},\mathbf{C\mu_{gh}}^T, \mathbf{P\mu_{gh}}^T]^T$). Distance d($\mathbf{x_{uv}}$, $\mathbf{x'_{uv}}$) is as defined earlier,

(4) $$d(\mathbf{x_{uv}}, \mathbf{x'_{uv}}) = sqE(z_{uv}, \mu_{gh}) + w_1 sqE(C_u, C\ \mu_{gh}) + w_2 sqE(P_v, P\ \mu_{gh})$$

Each co-cluster hence has a customer prototype, a product prototype and a mean preference value. Bregman block average co-clustering identifies uniform blocks with similar values as co-clusters. The aim in this case is to find co-clusters of customers and products with similar preference values, customer characteristics and product attributes. In the ideal case, all the customers and products in a co-cluster have the same attributes and a constant preference value. Any deviations from this are the effects of noise. The generative noise model assumed by this approach is a Gaussian mixture model consisting of k*l components. Each component is a multivariate Gaussian with a diagonal covariance matrix, where the diagonal elements are 1,1/w1,1/w1....,1/w2,1/w2...etc. Minimizing the objective function in equation (3) is equivalent to maximizing the log likelihood of the underlying generative model. If interpreted in this way, attribute-based co-clustering is a parametric approach since it makes an assumption of the probability model and tries to estimate the parameters of the model from the given data.

A simple iterative algorithm can be used to find a co-clustering (ρ, γ), that minimizes the objective function in equation (3). The objective function is the sum of the error between the original and predicted vectors over all the customer-product pairs in the matrix. It can hence be expressed as a sum of row/column errors. If row (customer) u is assigned to row cluster g. (ρ(u)=g), the row error is

(5) $$E_u(g) = \sum_{h=1}^{l} \sum_{v:\gamma(v)=h} d(x_{uv}, x_{uv}')$$

The best choice of the row cluster assignment for row u is the g that has the smallest error

(6) ρ^{new}(u)=argmin_g E_u(g)

A similar approach is used to assign columns to column clusters. Assigning each row and column to the row/column cluster with the smallest error reduces the objective function. We present the resulting algorithm in detail in Appendix G. The weights w1 and w2 are determined by cross-validation. Attribute based co-clustering is actually a generalization of the Bregman bock average co-clustering algorithm such that, if the weights w1 and w2 are set to zero, the algorithm is then equivalent to the Bregman block average co-clustering.

Convert attribute-based co-clusters into recommendations. The co-cluster statistics, namely the prototype customer and product attributes and the mean preference value, are used to make recommendations for customers or products. When product recommendations are made for a customer c with characteristics $\mathbf{C_{new}}$, the customer has to first be assigned to a row cluster. The customer is assigned to the row cluster g that has the closest customer prototype $\mathbf{C\mu_g}$ to $\mathbf{C_{new}}$.

(7) $$\rho(c) = \text{argmin}_g \text{sqE}(C_{new}, C\,\mu_g)$$

The co-cluster statistics, namely the prototype customer and product attributes and the mean preference value, are used to make recommendations for customers or products. From among the l column clusters in the assigned row cluster g, the column cluster h with the highest mean preference value μ_{gh} would be the cluster consisting of the set of products that the new customer is most likely to purchase. All the products in the row cluster g and column cluster h are recommended to the new customer. A similar approach is used to recommend customers, given a new product. The product is first assigned to the closest column cluster h. From among the row clusters in the column cluster h, the cluster with the highest mean cell value is selected and the new product is recommended to the customers in this co-cluster.

Latent Class Segmentation

In this research, we compare the performance of attribute-based co-clustering with two variations of a latent class segmentation recommendation system approach: product segmentation and customer segmentation. The product segmentation approach (Godfrey, McAlister, Saar-Tsechansky 2007) identifies product segments such that products in a segment attract the same type of customers. The approach begins with data for a sample of customers' choices across many products. The problem is made tractable by reducing the large number of individual customers to a few managerially relevant customer-types. Using the attraction model (McAlister, George, Chien 2007), for each product, we determine the relative strength with which that product attracts the different customer-types. We then identify a finite number of latent product segments such that products in a segment attract the same customer-types. Finally, in a posterior analysis, we probabilistically assign products to segments.

The product segmentation approach parallels latent class approaches that identify customer segments such that customers in a segment prefer the same type of products. In particular, this methodology follows the approach first presented by Kamakura and Russell (1989) to identify customer segments based on customers' preferences for product-types. The customer segmentation approach begins with data for a sample of customers' choices across many products. In this case, the problem is made tractable by reducing the large number of products to a few managerially relevant product-types. Using a multinomial logit choice model, for each customer, we determine the relative strength with which that customer prefers the different product-types. Using latent class analysis, the customer segmentation approach identifies a finite number of latent customer segments and, in a posterior analysis, probabilistically assigns customers to segments. Because latent class customer segmentation is so widely applied, we provide a detailed presentation of the product segmentation approach here and refer the reader to Kamakura and Russell (1989) for further detail on the parallel customer segmentation approach. A graphical representation of product segmentation and customer segmentation, relative to the proposed attribute-based co-clustering approach is illustrated in Figure 1.

Identify latent segments. To determine the relative strength with which a product attracts different customer-types, we apply McAlister, George and Chien's (2007) attraction model. This model represents the strength with which a product attracts each customer-type as the conditional probability that a given purchase of the product, p, was made by a particular customer-type, C, on a particular transaction, t, $prob(C,t \mid p)$. Specifically, given a set of customer-types, C =1,2,…, N_C and a set of individual products, p = 1,2,…,n_P, we define, for the randomly selected t^{th} transaction on which product p was chosen, the probability that the purchase was made by a customer of type C as:

$$prob(C,t \mid p) = \exp\{\alpha_{p,C,t}\} \Big/ \sum_{C=1}^{N_C} \exp\{\alpha_{p,C,t}\} \tag{8}$$

Given a set of M observable customer, product, and market environment variables that influence the strength with which product p attracts different customer-types, the deterministic component of the strength with which product p attracts a customer of type C on the t^{th} transaction is calculated as:

$$\alpha_{p,C,t} = \sum_{m=1}^{M} w_m x_{p,C,t} \tag{9}$$

where $x_{p,C,t}$ is the observed value of characteristic m for customer-type C and product p on the t^{th} transaction, and w_m is the attraction weight of characteristic m. The calculated probabilities represent a product's probability of attracting each customer-type on the t^{th} transaction. We refer to this set of probabilities as a product's "customer mix".

We define product segments by identifying products that have the same customer mix. That is, we allow for heterogeneity in the strength with which products attract different customer-types. We group together products using a mixture model that combines latent class analysis with the attraction model. Specifically, we assume there exists a finite number of product segments N_Π and define, for any given purchase of product p in product segment Π, the probability that the purchase was made by customer-type C on the t[th] transaction as:

(10)
$$prob(C,t \mid p\varepsilon\Pi) = Q_\Pi * prob(C,t \mid p)$$

where $Q_\Pi = \exp\{\theta_\Pi\} \Big/ \sum_{\Pi=1}^{N_\Pi} \exp\{\theta_\Pi\}$ is the unconditional probability that a given product p is included in product segment Π, and θ_Π is the estimated product segment size parameter. We estimate the model using maximum likelihood procedures to obtain estimates of the attraction weights for each product segment and the size of each product segment. Letting H_p be the collection of all transactions in which product p was chosen, the likelihood function is:

(11)
$$L(H_p) = \sum_{\Pi=1}^{N_\Pi} Q_\Pi * L(H_p \mid \Pi)$$

where $L(H_p \mid \Pi) = \prod_{C=1}^{N_C} \prod_{t=1}^{T} prob(C,t \mid p\varepsilon\Pi)$.

Finally, in a posterior analysis, we probabilistically assign each product to a product segment such that items assigned to a product segment attract the same customer-types. Specifically, we employ a Bayesian calculation to compute the probability that product p is included in product segment Π and assign each product to the product segment for which it has the highest inclusion probability. The segment assignment probabilities are calculated as:

(12)
$$prob(p\varepsilon\Pi \mid H_p) = L(H_p \mid \Pi) * Q_\Pi \Big/ \sum_{\Pi=1}^{N_\Pi} [L(H_p \mid \Pi) * Q_\Pi]$$

Convert latent segments into recommendations. When we apply latent class segmentation to recommend of a segment of products, Π, for a particular customer-type, C_0, to purchase, we use the conditional probabilities $prob(\Pi \mid C_0)$. The conditional recommendation probabilities are calculated as $prob(\Pi \mid C_0) = prob(C_0 \mid \Pi) * prob(\Pi) / prob(C_0)$, where $prob(C_0 \mid \Pi)$ and $prob(\Pi)$ are calculated from parameters estimated by the model and $prob(C_0)$ is observed from the sample of customers. As such, we use the probability that a particular kind of product was chosen (a product from segment Π) given that a customer of type C_0 did the choosing to recommend products for a customer to purchase.

When we apply latent class segmentation to recommend a segment of customers, χ, for a particular product-type, P_0, to target, we use the conditional probabilities $prob(\chi \mid P_0)$. The conditional recommendation probabilities are calculated as $prob(\chi \mid P_0) = prob(P_0 \mid \chi) * prob(\chi) / prob(P_0)$, where $prob(P_0 \mid \chi)$ and $prob(\chi)$ are calculated from parameters estimated by the model and $prob(P_0)$ is observed from the sample of products. As such, we use the probability that a particular kind of customer did the choosing (a customer from segment χ) given that a product of type P_0 was chosen to recommend customers for a product to target.

Nearest Neighbor Collaborative Filtering

In this research, we also compare the performance of the proposed attribute-based co-clustering approach with a well-established and widely applied recommendation system, nearest neighbor collaborative filtering. Collaborative filtering recommendation systems work by collecting customer ratings or choices for products or services and identifying a set of customers who share the same information needs or the same tastes. The objective of the collaborative filtering system is to then provide personalized product recommendations. While extensive research has evaluated the performance of collaborative filtering algorithms using explicit data, little work has evaluated collaborative filtering using implicit ratings (Herlocker et al. 1999).

Collaborative filtering recommendation systems generally perform well at predicting products that match a customer's interests or tastes, however these systems are not well suited to providing information on the factors that drive those predictions because they do not incorporate any information on customer characteristics or product attributes (Herlocker et al. 1999). The primary advantage of collaborative filtering is that it can develop recommendations on tacit qualities, beyond customer characteristics and product attributes, such as quality and taste.

Identify nearest neighbors. In collaborative filtering, the problem space is formulated as a matrix of customers and products with each cell representing a customer's rating or choice of a specific product. The most prevalent algorithms used in collaborative filtering are referred to as neighborhood-based methods. In neighborhood-based methods, a subset of appropriate customers is chosen, based on their similarity to the target customer and a weighted aggregate of their ratings is used to generate predictions for the target customer. The recommendation procedure essentially comprises three steps. First the approach calculates a weight for all customers with respect to the target customer. Next, the approach selects a subset of customers, or "nearest neighbors" to use in making recommendations. Finally, the approach normalizes the ratings or choices and develops a recommendation from a weighted combination of selected neighbors' ratings or choices. Within specific systems, these steps may overlap or the order may vary and different systems may use different measures to calculate customers' similarity. For example, some systems use a Pearson correlation to weigh customers' similarity and then compute final recommendations by performing a weighted average of deviations from the neighbor's mean. Other approaches calculate the Euclidian distance between customers' prior ratings or choices to compute similarity. We employed the latter approach here.

Convert nearest neighbors into recommendations. For each product recommended, the highest ranking or closest neighbors are used to compute a recommendation. That is, the set of customers form the customer's neighborhood for that item. All customers in the database are examined as potential neighbors for a customer. After the approach assigns similarity weights to customers in the database, it then determines which other customer's ratings or choice data will be used in the computation of a recommendation for the target customer. A common approach to selecting the neighborhood is to specify an arbitrary number of the nearest neighbors or a maximum distance. In this research, we employ a fixed neighborhood size.

Once the neighborhood has been selected, the ratings or choices from those neighbors are employed to compute a recommendation. The basic way to combine all of the neighbors' ratings or choices into a recommendation is to compute an average of the ratings. In cases where the data capture binary choices, recommendations are developed using a rule such as recommending a product if at least half of the neighbors purchased that item.

Compare Approaches

Given that attribute-based co-clustering, latent class product / customer segmentation, and nearest neighbor collaborative filtering represent three different recommendation system approaches that may be differentially impacted by the information used and the features of the data sample, we evaluate the relative performance of the three approaches by comparing the quality of their recommendations in two settings. The first setting investigates each approach's effectiveness at recommending a group of products for a customer to purchase. To evaluate the quality of the recommendations for each approach, we compare the set of products each approach recommends for a withheld customer to purchase with the set of products that customer actually purchased. The second setting investigates each approach's effectiveness at recommending a group of customers for a product to target. Although recommending customers for a product to target has not been explored in the recommendation system literature, it represents another useful application of recommendation systems. Further, applying recommendation systems in this second test allows us to evaluate the performance of the three recommendation systems using data with different dimensions. That is, in this case, the products represent the rows and the customers represent the columns of the data matrix. In this test, we compare the set of customers each approach recommends for a withheld product to target with the set of customers who actually purchased that product. In both tests, we evaluate each approach using the leave-one-out variation of the n-fold bootstrapping technique (see Mitchell 1997). In a data set with N observations, this technique involves applying the approach N separate times on all of the data except for one observation (i.e., estimate the model with N-1 observations) and then making a prediction for the withheld observation.

First test: Recommend products for a customer to purchase. In the first test, we apply attribute-based co-clustering, product segmentation, and collaborative filtering to recommend a group of products for a customer to purchase. Note that, since Godfrey, McAlister and Saar-Tsechansky (2007) find that the product segmentation approach performs better than the customer segmentation approach at recommending a set of products for a customer to purchase, we apply only the product segmentation approach in this test. For attribute-based co-clustering and product segmentation, we withhold a customer and disregard all information about the customer except the customer-type, estimate the model using the purchase histories of all other customers in the data set, and then identify a group of products to recommend to the withheld customer. Because nearest neighbor collaborative filtering relies solely on observed choices to make recommendations, this approach requires some partial knowledge about the customer for which a recommendation is required. As such we follow a different procedure to make recommendations. For collaborative filtering, we withhold a customer and randomly select a subset of the products chosen by that customer. We give the algorithm that subset of products to use in identifying the nearest neighborhood of customers whose choices are used to recommend other products. For each approach, we repeat the process for each customer in the data set and calculate the quality of the recommendations for each customer by comparing the recommended products with the products the customer actually purchased.

Second test: Recommend customers for a product to target. In the second test, we apply attribute-based co-clustering, customer segmentation, and collaborative filtering to recommend a group of customers for a product to target. Note in this case that, since Godfrey, McAlister and Saar-Tsechansky (2007) find that the customer segmentation approach performs better than the product segmentation approach at recommending a set of customers for a product to target, we apply only the customer segmentation approach in this test. For attribute-based co-clustering and customer segmentation, we withhold a product and disregard all information about the product except its product-type, estimate the model using the purchases of all other products in the data set, and then identify a group of customers to whom the withheld product should be targeted. Again, since collaborative filtering relies solely on observed choices to make recommendations, we must follow a different procedure and allow the procedure to have some partial knowledge about the product for which a recommendation is required. For collaborative filtering, we withhold a product and randomly select a subset of the customers who chose that product. We give the algorithm that subset of customers to use in identifying the nearest neighborhood of products. The customers who chose this product are used to recommend customers for the withheld product to target. We repeat the process for each product in the data set and calculate the quality of the recommendations for each product by comparing the recommended customers with the customers who actually purchased the product.

Metrics used to compare quality of recommendations. To evaluate the performance of the three recommendation system approaches in these two tests, we compare the recommendations made by each approach in terms of three objective metrics: precision, recall and the F metric. Precision and recall are among the most popular metrics for evaluating and comparing recommendation systems and are particularly relevant metrics for applications using binary data (Herlocker et al. 2004). Precision represents the probability that a recommended item (e.g., product or customer) is relevant and is calculated as the ratio of the number of recommended items actually chosen to the total number of items recommended for the withheld user. Recall represents the probability that a relevant item is recommended and is calculated as the ratio of the number of recommended items actually chosen to the total number of relevant items for the withheld user. These metrics are often referred to as the "hit rate" in prior comparisons of recommendation systems in the marketing literature. Recall and precision involve trade-offs that are related to the number of items recommended to the user. For example, when more items are recommended, recall tends to increase and precision tends to decrease. Thus, to fully describe the performance of each approach by considering recall and precision together and to account for the fact that the number of items recommended varies across approaches, we compare the three approaches in terms of a third metric, F, where F is calculated as [2 * Precision * Recall] / [Precision + Recall].

CHAPTER 9: EMPIRICAL APPLICATION

To evaluate the relative performance of the proposed attribute-based co-clustering approach, latent class product / customer segmentation, and nearest neighbor collaborative filtering recommendation systems, we apply the three approaches to make recommendations in a service context. Specifically, we examine the elective courses chosen by MBA students in the business school at a large southwestern university. As a service provider, the business school offers a large number of elective courses to meet the needs of different types of students. Specifically, the business school offers a range of elective courses across multiple departments (accounting, finance, management, management of information systems, and marketing) to meet the needs of students obtaining an MBA degree to pursue careers in investment banking, corporate finance, technology management, general management, brand management, and consulting, among other fields. As such, the business school is a service provider for which the products (services) are courses that can be defined by the departments that offer those courses; and for which the customers are students that can be defined by the careers the students want to pursue.

The problem of aligning elective courses offered by the business school and students having different career objectives can be viewed from two perspectives. First, one can consider the perspective of students who must decide which elective courses to choose from the large course offering in preparation for a particular career. To assist a student pursuing a particular career in choosing which courses to take, an academic advisor needs to know what set of courses to recommend to that student. Thus, the objective of the first test in which we compare the three approaches is to identify the set of courses that should be recommended for a particular student to take. Alternatively, one can consider the perspective of the business school administrators and professors who must decide which students among the large student body to target in designing the program curriculum or managing course enrollment. For example, to assist a professor in a particular department in attracting students to enroll in her course, the professor needs to know to what set of students the course should be targeted. Thus, the objective of the second test in which we compare the three approaches is to identify the set of students that should be recommended for a particular course to target.

Description of Data

The data used in these applications of the three recommendation system approaches includes two sets of information. The first data set describes the set of 32 product (course) alternatives and identifies which students enrolled in each course. The data set comprises course enrollment data for elective MBA courses offered during the 1998-99 and 1999-2000 academic years as reported by the university's MBA program office. As such, the course enrollment information represents an implicit, binary data set where the rows are students and the columns are courses and each cell indicates whether a particular student enrolled in a given course. Compulsory courses were omitted from the analysis because these courses are required of all students and, therefore, have no observable variation in attraction across students. We find variation among elective course choices because students in this MBA program are not required to declare a concentration, but rather can choose courses offered by any department based on their interests, strengths and perspectives on how best to prepare for a particular career. Based on input from MBA program administrators on factors that might help explain courses' attraction for different students, the course attribute included in the analysis is the department in which the course is offered. Courses are offered by five different departments: (1) Accounting, (2) Finance, (3) Management of Information Systems, (4) Management, and (5) Marketing. A summary of the course characteristic is presented in Table 11.

We use the observations of which students enrolled in each course in the first data set to merge the information with a second data set. The second data set describes the sample of 326 customers (students) who graduated from the MBA program in 2000. This information was derived from a survey completed by all students upon graduation. The student characteristic included in the model was also based on input from MBA program administrators and helps explain students' career orientation. The characteristic included is the first job the student took after graduation: (1) Investment Banker, (2) Corporate Finance, (3) Technology Manager, (4) General Manager, (5) Brand Manager, (6) Consultant, and (7) Other. A summary of the student characteristics is presented in Table 11.

TABLE 11
RECOMMENDATION SYSTEM DATA CHARACTERISTICS

Products (Courses)		**Customers (Students)**	
Product-type sample shares		Customer-type sample shares	
Accounting	9%	Investment banker	15%
Finance	28%	Corporate finance	9%
Management of information systems	22%	IT manager	10%
Management	32%	General manager	12%
Marketing	9%	Brand manager	18%
		Consultant	19%
		Other	17%
Total # of courses	32	Total # of students	326
Mean # courses taken per student	8	Mean # students enrolled per course	82
Average density of course choices	0.25	Average density of student enrolment	0.25

Results

To evaluate the relative performance of the three recommendation approaches, we compare the quality of the recommendations made by each approach in two different tests. In the first test, we apply each approach to recommend a set of courses for a withheld student to take. In the second test, we apply each approach to recommend a set of students for a withheld course to target. To compare results, we compare the mean recall, precision, and F metric for each approach using paired sample t-tests.

Comparison of quality of product recommendations. Table 12 presents a comparison of the quality of recommendations made by each approach for Test 1, where we apply each of the three methods to make course recommendations. The results for Test 1 indicate that collaborative filtering generally performs better than the other approaches. In terms of precision, when collaborative filtering is given 2 randomly selected courses, the recommendations are significantly better than the product segmentation recommendations (.51 vs. .27, $p < .01$). When the number of courses given to collaborative filtering is increased to 4 randomly selected courses, the precision of the recommendations decreases, however the recommendations are still significantly better than the product segmentation recommendations (.47 vs. .27, $p < .01$).

TABLE 12
RELATIVE QUALITY OF RECOMMENDATIONS OF COURSES FOR A STUDENT TO TAKE

Approach	Recall	Precision	F
Collaborative Filtering (given 2 courses)	.38 [c, d]	.57 [b, c, d]	.43 [b, c, d]
Collaborative Filtering (given 4 courses)	.38 [c, d]	.47 [a, c, d]	.39 [a, c, d]
Attribute-based co-clustering	.41 [a, b, d]	.37 [a, b, d]	.37 [a, b, d]
Product Segmentation	.50 [a, b, c]	.27 [a, b, c]	.33 [a, b, c]

Notes:
[a] indicates that result is significantly different from Collaborative Filtering (given 2 courses) at $p < .05$
[b] indicates that result is significantly different from Collaborative Filtering (given 4 courses) at $p < .05$
[c] indicates that result is significantly different from Proposed attribute-based co-clustering (base) at $p < .05$
[d] indicates that result is significantly different from Product Segmentation (base) at $p < .05$

Two features of the data set in this application can help explain why collaborative filtering recommendations have higher precision than product segmentation recommendations. First, product segmentation, which is a regression-based method, tends to perform better when the sample is large relative to the number of alternatives in the choice set. In this test, product segmentation identifies the segment of courses to recommend by comparing attraction patterns across a set of 326 students (excluding the withheld student) for a sample of 32 courses. As such, the sample of 32 courses is much smaller than the dimensionality of the set of 326 students. Thus, the small size of the sample of courses negatively impacts the quality of recommendations made by product segmentation. Second, as the number of choice alternatives increases, the quality of collaborative filtering recommendations decreases because the prediction task increases in difficulty with more alternatives. In this test, collaborative filtering identifies the neighborhood of courses to recommend by comparing the courses the withheld student took with the patterns of 32 courses chosen by the remaining sample of 325 students. As such, the dimensionality of the choice set of 32 courses for collaborative filtering is much smaller than the dimensionality of the choice set of 326 students for the product segmentation approach. The smaller dimensionality makes it easier for collaborative filtering to identify nearest neighbors. Thus, the smaller dimensionality of the choice set positively impacts the quality of the collaborative filtering recommendations.

Contrary to the general trend of results for Test 1, collaborative filtering performs worse than the other approaches in one respect. In terms of recall, when collaborative filtering is given 2 courses, the recommendations are significantly worse than the product segmentation recommendations (.38 vs. .50, $p < .01$). Collaborative filtering recommendations also have lower recall than product segmentation recommendations when the number of courses given to collaborative filtering is increased to 4 courses (.38 vs. .50, $p < .01$). This exception to the general trend of relative results is not entirely surprising. The recall of nearest neighbor collaborative filtering recommendations tends to be lower than the recall of the other approaches because, by their structure, the other approaches recommend a set of many courses, while nearest neighbor collaborative filtering examines and recommends each course individually. As such, this collaborative filtering approach tends to recommend fewer courses and is likely to exclude a number of courses the withheld student actually took, which results in lower recall. In contrast, methods that recommend segments, such as product segmentation, or clusters, such as proposed attribute-based co-clustering tend to recommend larger sets of items and, thus, have a lower chance of missing courses that the student actually took. Because the recall of the collaborative filtering recommendations is so low in this test, it performs worse than the other approaches on this metric.

Further comparison of the recall and precision of the recommendations made by each approach indicates that attribute-based co-clustering recommendations tend to be the most stable and the quality of the recommendations falls between the other two approaches. In terms of precision, attribute-based co-clustering recommendations were significantly worse than collaborative filtering (.37 vs. .57, $p < .01$ given 2 courses, .37 vs. .47, $p < .01$ given 4 courses) but were significantly better than product segmentation (.37 vs. .27, $p < .01$). In terms of recall, attribute-based co-clustering recommendations were significantly better than collaborative filtering (.42 vs. .38, $p < .01$ given 2 courses and given 4 courses) but were significantly worse than product segmentation (.42 vs. .50, $p < .01$). These results can be explained by the fact that, because attribute-based co-clustering identifies clusters of courses to recommend by analyzing students and courses simultaneously, the sample size and dimensionality of the choice set is the same in both tests, and thus the performance in both tests is similar. That is, attribute-based co-clustering is not significantly disadvantaged by the low sample size that impacts product segmentation but is not significantly advantaged by the small dimensionality of the choice set that impacts collaborative filtering.

Some of the results presented above indicate that the quality of collaborative filtering recommendations, relative to the other two approaches, depends on the amount of information available for the withheld student. That is, the differences in the quality of the recommendations across approaches changes in some cases because collaborative filtering recall tends to increase and the precision tends to decrease when that approach is given a higher number of courses the withheld student took. To gain a better understanding of the impact of the density of data for the withheld student on the quality of collaborative filtering recommendations, we further investigate the impact of this data feature. Additionally, we examine the impact of the density of data for the withheld student on the quality of recommendations for attribute-based co-clustering and latent class product / customer segmentation. In the original comparisons across approaches, which we refer to as the base case, attribute-based co-clustering and latent class product / customer segmentation are not given any information on courses the withheld student took. As such, the base case essentially represents the quality of recommendations to a new student. To compare the quality of recommendations for these two approaches in the case of an existing student, where additional information is available, we give the attribute-based co-clustering and latent class product / customer segmentation approaches the same information that is given to collaborative filtering and evaluate the impact of the density of that additional data.

Impact of data density on quality of product recommendations. The results in Table 13 illustrate the impact of giving additional observations of courses the withheld student took on each of the three approaches in terms of recall and precision. The results indicate that giving product segmentation and attribute-based co-clustering additional information on courses taken by the withheld student generally has no significant effect on recall and precision. Compared to the base case for these two approaches, differences in recall and precision when the known courses are added are not statistically significant. The additional information does not impact the recommendations of these approaches because product segmentation and attribute-based co-clustering develop their recommendations primarily based on other information, besides the known courses for the withheld student. Specifically, product segmentation develops its recommendations by estimating a latent class model that uses information on the students' career type. Attribute-based co-clustering performs a clustering algorithm that uses information on the students' career type and the courses' department as well as an a priori specified number of course clusters and student clusters. These two approaches only use the information on the known courses in the final calculation of which segment or cluster to recommend. Since the model estimation is not impacted by the addition of information on the known courses and, because the additional information on the known courses represents just one type of information used by these two approaches, the additional information does not significantly impact the recommendations.

One set of results runs contrary to this expectation. Adding known courses for the withheld student to product segmentation has a significant negative impact on recall in this test. Compared to the base case, giving two courses significantly decreases the recall of the recommendations (.50 vs. .45, $p = .01$). Adding four courses also decreases recall compared to the base case (.50 vs. .46, $p = .02$). This negative impact can be attributed to known courses that are outliers. That is, when the course observations added to the product segmentation's calculation of the probability of recommending a particular course segment are representative of the types of courses typically taken by the withheld student, the additional information reinforces the recommendation probabilities for each course segment. However, when the additional course observations are not representative of the types of courses typically taken by the withheld student, the additional information changes the course segment recommendation probabilities such that the recommendations have lower recall.

TABLE 13
IMPACT OF DATA DENSITY ON QUALITY OF RECOMMENDATIONS OF COURSES FOR A STUDENT TO TAKE

	Product Segmentation		**Attribute-based Co-clustering**		**Collaborative Filtering**	
	Recall	Precision	Recall	Precision	Recall	Precision
Base (Given no courses)	.50 [b, c]	.27	.41	.37		
Given 2 courses	.45 [a]	.27 [c]	.41	.37	.38	.57 [c]
Given 4 courses	.46 [a]	.30 [b]	.40	.37	.38	.47 [b]

Notes:
[a] indicates that result is significantly different from Base (Given no courses) case at $p < .05$
[b] indicates that result is significantly different from Given 2 courses case at $p < .05$
[c] indicates that result is significantly different from Given 4 courses case at $p < .05$

In contrast to product segmentation and attribute-based co-clustering, which primarily use information on course attributes and student characteristics to develop recommendations, collaborative filtering relies solely on the observations of courses taken by the withheld student and those selected by its closest neighbors (which is also determined by the information available on the withheld item) to develop recommendations. As such, we expect the number of observations given to the algorithm to impact the quality of the recommendations. When little information is provided to describe the withheld student, no close neighborhood is identified and the neighbors are likely to have fewer common preferences or choices. Thus, very few courses are recommended and recall is lower. Precision, however, is high because the items recommended are more likely to be simply “popular” items rather than choices that are specific to the customers in the neighborhood. In contrast, when collaborative filtering is given more information in terms of additional courses the withheld student took, the algorithm is able to identify more of courses that are similar to the courses taken by the withheld student. As more information is provided on the withheld item, the neighbors are likely to have more common preferences/choices. Thus, more items are recommended the recall of the recommendation increases. Giving more courses the withheld student took to collaborative filtering may negatively impact precision, however. When given additional information, the algorithm is better able to identify a neighborhood of courses that includes courses other than popular courses that all students tend to take. Because collaborative filtering recommends only the choices made by these neighbors, the precision of the recommended neighborhood of courses is likely to decrease. The results for Test 1 generally support this expectation. The results in Table 13 indicate that giving collaborative filtering additional information on courses taken by the withheld student has no significant effect on recall, but significantly decreases precision. Compared to the case where the approach is given 2 courses, precision is significantly lower when the approach is given 4 courses (.57 vs. .47, $p < .01$). As such, we find that the amount of available information on the withheld student, in terms of the number of courses the student took that are given to collaborative filtering impacts the quality of recommendations for this approach.

Impact of predictive variables versus mere choices on product recommendations. A more fundamental explanation for the higher quality of collaborative filtering recommendations relative to product segmentation and attribute-based co-clustering relates to the inherent ability of collaborative filtering to identify underlying patterns in choices when explanatory variables are not available or are not good predictors of behavior. In this first test, product segmentation develops its recommendations using information on the students' career type. As such, the quality of the product segmentation recommendations depends on how well these variables predict the courses students take. Attribute-based co-clustering uses information on both the students' career type and the courses' department. The predefined student-types impact the results of attribute-based co-clustering in this test because this approach assumes that students with similar attributes (career types) make similar choices of courses. If this assumption is not entirely true, this will negatively impact the quality of attribute-based co-clustering recommendations. In contrast, collaborative filtering is independent of the student career and course department information; this approach performs well when past behavior, rather than (available) explanatory variables, is largely indicative of future choice.

One objective measure of the amount of information provided by these variables is the normalized mutual information (NMI). When we calculate the NMI of our explanatory variables, we find that the NMI of the student careers is .02 while the NMI of the course departments is .23. As such, based on this measure, the student careers appear to provide less information than the course departments. Since product segmentation relies on the student careers to develop its recommendations, and these variables provide low information, the product segmentation recommendations are generally worse than the recommendations made by collaborative filtering, which does not use this information and is able to identify underlying patterns. In addition to the student careers, attribute-based co-clustering uses the course departments, which provide relatively more information. Because attribute-based co-clustering assigns higher weights to attributes that are more predictive of choice, this approach places more weight on the course departments and less weight on the student careers. As such, the quality of the attribute-based co-clustering recommendations is generally better than the product segmentation recommendations. However, because collaborative filtering is able to identify additional underlying patterns not identified by proposed attribute-based co-clustering, the quality of the collaborative filtering recommendations tends to be better than attribute-based co-clustering.

Comparison of quality of customer recommendations. Table 14 presents a comparison of the quality of recommendations made by each approach for Test 2 in which we apply the each of three methods to recommend a set of students for a withheld course to target. In contrast to the results for Test 1, the results for Test 2 indicate that collaborative filtering generally performs worse than the other approaches in terms of recall and precision. In terms of recall, when collaborative filtering is given 2 students, recall is significantly worse than customer segmentation (.29 vs. .60, $p < .01$). When the number of students given to collaborative filtering is increased to 10 students, differences between collaborative filtering and customer segmentation are not statistically significant. In terms of precision, when collaborative filtering is given 2 students, differences between collaborative filtering and customer segmentation are not statistically significant. However, when the number of students given to collaborative filtering is increased to 10 students, precision is significantly worse than customer segmentation (.31 vs. .37, p = .04).

TABLE 14
RELATIVE QUALITY OF RECOMMENDATIONS OF STUDENTS FOR A COURSE TO TARGET

Approach	Recall	Precision	F
Customer Segmentation	.60 [d]	.37 [c]	.40 [d]
Attribute-based co-clustering	.56 [d]	.35 [c]	.38 [d]
Collaborative Filtering (given 10 students)	.52 [d]	.31 [a, b]	.35 [d]
Collaborative Filtering (given 2 students)	.29 [a, b, c]	.34	.23 [a, b, c]

Notes:
[a] indicates that result is significantly different from Customer Segmentation (base) at $p < .05$
[b] indicates that result is significantly different from Proposed attribute-based co-clustering (base) at $p < .05$
[c] indicates that result is significantly different from Collaborative Filtering (given 10 students) at $p < .05$
[d] indicates that result is significantly different from Collaborative Filtering (given 2 students) at $p < .05$

We can explain why collaborative filtering generally performs worse than customer segmentation in this test by again considering the sample size and dimensionality of the choice set as applied in each approach. Since customer segmentation is a regression-based method, it tends to perform better when the sample is large relative to the set of choice alternatives. In this test, customer segmentation identifies the segment of students to recommend by comparing choice patterns across a set of 32 courses (excluding the withheld course) for a sample of 326 students. As such, the sample of 326 students is much larger than the dimensionality of the set of 32 courses. Thus, this large sample size positively impacts the quality of recommendations made by customer segmentation. In this test, collaborative filtering identifies the neighborhood of students to recommend by comparing the students who took the withheld course with the patterns of 326 students who chose the remaining sample of 31 courses. As such, the dimensionality of the choice set of 326 students is very large compared to the dimensionality of the choice set for the product segmentation approach. The large dimensionality makes it difficult for collaborative filtering to identify nearest neighbors and, thus, this negatively impacts the quality of the collaborative filtering recommendations, relative to the customer segmentation recommendations. The finding that some differences between customer segmentation and collaborative filtering are statistically significant while others are not can be attributed to the sensitivity of collaborative filtering to the density of information for the withheld course.

Comparing results for Test 2 in Table 14 also indicates that attribute-based co-clustering generally performs about the same as customer segmentation in terms of recall and precision. In this test, differences in recall and precision for attribute-based co-clustering and customer segmentation are not significant. Results also indicate that attribute-based co-clustering generally performs better than collaborative filtering. In terms of recall, when collaborative filtering is given 2 students, recall is significantly worse than attribute-based co-clustering (.29 vs. .56, $p < .01$). When the number of students given to collaborative filtering is increased to 10 students, differences between collaborative filtering and attribute-based co-clustering are not statistically significant. In terms of precision, when collaborative filtering is given 2 students, differences between collaborative filtering and attribute-based co-clustering are not statistically significant. However, when the number of students given to collaborative filtering is increased to 10 students, precision is significantly worse than attribute-based co-clustering (.31 vs. .35, p = .04). Attribute-based co-clustering recommendations have higher precision and recall than collaborative filtering because attribute-based co-clustering is not significantly impacted by the high dimensionality of the choice set that impairs collaborative filtering from making quality recommendations in this test. That is, since attribute-based co-clustering identifies clusters of courses to recommend by clustering students and courses simultaneously, the sample size and dimensionality of the choice set have less impact on this approach and it is able to make better recommendations than collaborative filtering. The fact that some differences between attribute-based co-clustering and collaborative filtering are statistically significant while others are not can again be attributed to the sensitivity of collaborative filtering to the density of information for the withheld course.

Some of the results presented above indicate that the quality of collaborative filtering recommendations, relative to the other two approaches, depends on the amount of information available for the withheld student. That is, the differences in the quality of the recommendations across approaches changes in some cases because collaborative filtering recall tends to increase and the precision tends to decrease when that approach is given a higher number of students who took the withheld course. To gain a better understanding of the impact of the density of data for the withheld course on the quality of collaborative filtering recommendations, we further investigate the impact of this data feature in the context of recommending students for a course to target. Additionally, we examine the impact of the density of data for the withheld course on the quality of recommendations for attribute-based co-clustering and latent class product /

customer segmentation in this context. In the original comparisons across approaches, which we refer to as the base case, attribute-based co-clustering and latent class product / customer segmentation are not given any information on students who took the withheld course. As such, the base case essentially represents the quality of recommendations for a new course. To compare the quality of recommendations for these two approaches in the case of an existing course, where additional information is available, we give the attribute-based co-clustering and latent class product / customer segmentation approaches the same information that is given to collaborative filtering and evaluate the impact of the density of that additional data.

Impact of data density on customer recommendations. The results in Table 15 illustrate the impact of giving additional observations of students who took the withheld course on each of the three approaches in terms of recall and precision. The results indicate that giving customer segmentation and attribute-based co-clustering additional information on students who took the withheld course has no significant effect on recall and precision. Compared to the base case for these approaches, differences in recall and precision when the known students are added are not statistically significant. As described earlier, the additional information does not impact the recommendations for these approaches because customer segmentation and attribute-based co-clustering develop their recommendations primarily based on other information, besides the known students for the withheld course. As such, we would not expect it to have a significant impact on the quality of the recommendations made by these two approaches.

TABLE 15
IMPACT OF DATA DENSITY ON QUALITY OF RECOMMENDATIONS OF STUDENTS FOR A COURSE TO TARGET

	Customer Segmentation		**Attribute-based Co-clustering**		**Collaborative Filtering**	
	Recall	Precision	Recall	Precision	Recall	Precision
Base (Given no students)	.60	.37	.56	.35		
Given 2 students	.55	.34 [c]	.48	.31	.29 [c]	.34
Given 10 students	.60	.38 [b]	.48	.33	.52 [b]	.31

Notes:
[a] indicates that result is significantly different from Base (Given no courses) case at $p < .05$
[b] indicates that result is significantly different from Given 2 students case at $p < .05$
[c] indicates that result is significantly different from Given 10 students case at $p < .05$

In contrast, because collaborative filtering relies solely on this information to develop recommendations, we expect the number of given students to impact the quality of the recommendations. As described above, giving additional information in terms of the number of students who took the enrolled course to collaborative filtering is likely to increase the recall and decrease the precision of recommendations made for this approach. The results for Test 2 generally support this expectation. The results in Table 15 indicate that giving collaborative filtering additional information on students who took the withheld course has no significant effect on precision, but significantly increases recall. Compared to the case where the approach is given 2 students, recall is significantly higher when the approach is given 10 students (.29 vs. .52, $p < .01$). As such, we find that the density of data for the withheld course, in terms of the number of students who took the withheld course that are given to collaborative filtering impacts the quality of recommendations for this approach.

Impact of predictive variables versus mere choices on customer recommendations. As with Test 1, a more fundamental explanation for the higher quality of customer segmentation and attribute-based co-clustering recommendations relative to collaborative filtering in this second test relates to the inherent ability of collaborative filtering to identify underlying patterns in choices when other explanatory variables are not available or are not good predictors of behavior. In this test, customer segmentation develops its recommendations using information on the courses' department while attribute-based co-clustering uses information on both the students' career type and the courses' department. Since customer segmentation relies on the course departments to develop its recommendations and these variables provide relatively high information (NMI = .23), the customer segmentation recommendations are generally better than the recommendations made by collaborative filtering, which do not use this information. In this case, the advantage collaborative filtering has in being able to identify underlying patterns is less relevant because the course departments are relatively good predictors of course choice patterns. In addition to the course departments, attribute-based co-clustering uses the student careers, which provide relatively little information (NMI = .02). Because attribute-based co-clustering assigns lower weights to attributes that are less predictive of choice, this approach places less weight on the student careers and more weight on the course departments. However, because customer segmentation does not use information on the student careers to any extent, the quality of the attribute-based co-clustering recommendations is generally worse than the customer segmentation recommendations.

CHAPTER 10: DISCUSSION

The results of our two tests show that, since nearest neighbor collaborative filtering does not use information on customer characteristics and product attributes, the quality of recommendations made by this approach does not depend on this information. The primary strength of this approach is that it is able to identify underlying choice patterns when customer and product information is not available or when the available information does not explain choices. We further show that nearest neighbor collaborative filtering recommendations are sensitive to the distribution of the data. Nearest neighbor collaborative filtering is sensitive to the dimensionality of the data set in terms of the number of alternatives in the choice set. When the dimensionality of the choice set is large, it is more difficult for this approach to identify a high quality neighborhood to recommend. Nearest neighbor collaborative filtering is also sensitive to the density of choices from the customer for whom recommendations are being made. Nearest neighbor collaborative filtering is sensitive to this feature because these choices comprise the only data used by this approach to make recommendations.

In contrast, the recommendations made by latent class product / customer segmentation are sensitive to the customer characteristics and product attributes included in the analysis. Since customer characteristics and product attributes represent the primary information used by latent class segmentation, the quality of the recommendations depends on the extent to which the characteristics and attributes explain the pattern of choices in the data. The primary strength of this approach is that, when the defined customer characteristics and product attributes do explain choice patterns, the parameters estimated by the approach provide insight into the factors that drive the choice patterns. Additionally, we show that the recommendations made by latent class segmentation are sensitive to the distribution of the data. Since latent class segmentation is a regression-based approach, it is better able to fit recommendations to the data when the sample size is large.

In general, we show that the proposed attribute-based co-clustering approach is the most stable recommendation system in that it is generally not as sensitive to the distribution of the data or the density of observations. Attribute-based co-clustering is not as sensitive to the distribution of the data set in terms of the sample size or the dimensionality of the choice set because it uses both dimensions of the information simultaneously. Attribute-based co-clustering is not as sensitive to the density of choices from the target customer because it uses other information, including customer characteristics and product attributes, in addition to choices, to make its recommendations. Further, because attribute-based co-clustering incorporates customer characteristics and product attributes, it is sensitive to the explanatory power of this information. However, because this approach incorporates customer characteristics and product attributes together, it is not as sensitive to cases where customer characteristics have relatively more or less information than product attributes. As such, we find that the quality of recommendations made by the proposed attribute-based co-clustering approach tends to be similar when the recommendation mechanism involves recommending products for customers to purchase or involves recommending customers for products to target, in terms of recall and precision.

In summary, our results suggest that latent class segmentation is likely to perform well in more mature customer transaction databases where there is a large sample of customers and high quality information on customer characteristics and product attributes is available. In contrast, collaborative filtering is likely to perform well in less mature customer transaction databases where the product choice set is relatively small and little or no quality information on customer characteristics or product attributes is available. The quality of collaborative filtering recommendations improves when there is a higher density of data for each customer in terms of a higher number of products chosen relative to the total choice set. Finally, while the proposed attribute-based co-clustering approach is likely to perform better when high quality information on customer characteristics and product attributes is available, this approach is generally the least sensitive to the distribution of the data. Because of the relative stability of this approach, attribute-based co-clustering presents an appealing methodology when the analyst does not wish to make significant trade-offs in terms of recall and precision.

Limitations and Future Research

This research contributes to our understanding of recommendation systems in three respects. First, we present a new recommendation system methodology, attribute-based co-clustering. The proposed attribute-based co-clustering approach extends prior co-clustering algorithms to incorporate information on customer characteristics and product attributes in the identification and recommendation of co-clusters. Second, we evaluate the relative performance of attribute-based co-clustering by comparing it to a related model-based approach, latent class segmentation, and a widely applied memory-based approach, nearest neighbor collaborative filtering. Finally, we investigate factors that impact the relative performance of the three types of recommendation systems. Specifically, we examine the impact of the quality of information on customer characteristics and product attributes and the impact of inherent features of the data set. We empirically evaluate the impact of these factors on the relative performance of the three approaches by comparing the quality of recommendations made by the three different approaches. We compare the quality of recommendations in two contexts relevant to retailers: recommending a set of products for a customer to purchase and recommending a set of customers for a product to target.

As with all comparisons of recommendation systems, our evaluation has limitations that present opportunities for future research. First, in this research, we identify a set of items to recommend (e.g., products) by identifying items that tend to be the most popular and tend to be similar to each other. As such, while it is highly likely that a target customer would like the products that are recommended, these recommendations do not present the customer with alternatives that the customer might like but are not similar to products the customer typically chooses. Thus, one direction for future research is to develop recommendation systems that identify products that are not similar to items frequently chosen by a particular customer, but are likely to be relevant to the customer.

Second, in this research, we evaluate the relative performance of the three different types of recommendation systems using three objective measures of recommendation quality: recall, precision and F metric. Because we use offline (i.e., secondary) data in our application of the three approaches, we are limited to this objective evaluation of the recommendations. However, these measures of recommendation quality do not reflect other aspects of recommendation quality such as the customer's satisfaction with the recommendations she received, whether the customer actually purchased a product based on the recommendations and, if a purchase transaction occurred, the customer's satisfaction with the product selected from the recommendation set. To capture such measures of recommendation system performance, future research should apply these approaches in experimental settings or field studies to evaluate the quality of recommendations along these alternate dimensions.

Finally, in our recommendation setting, all products (i.e., courses) and customers (i.e., students) are considered to have equal value to the service provider. In actual retail settings, however, products may have significant variation in value to the retailer. For example, some products have higher margins than do other products for a retailer. Similarly, some customers have higher lifetime value for a retailer than do other customers. As such, it would be valuable to extend the approaches presented here to allow the analyst to recommend a set of products for a customer to purchase that includes higher margin items or to identify a set of customers for a product to target that includes customers with higher lifetime value.

APPENDIX A
COMPARISON OF PRODUCT SEGMENTATION APPROACH AND CUSTOMER SEGMENTATION APPROACH SPECIFICATION AND ESTIMATION

		Product Segmentation Approach	Customer Segmentation Approach
	Notation	c = Index of individual customers n_c = Total number of individual customers C = Index of customer-types N_C = Total number of customer-types Π = Index of product segments N_Π = Total number of product segments $t_{p,C}$ = Index of all transactions including product p that were made by customers of type C (sometimes represented without subscripts to ease reading) $T_{p,C}$ = Total number of transactions including product p that were made by customers of type C	p = Index of individual products n_p = Total number of individual products P = Index of product-types N_P = Total number of product-types χ = Index of customer segments N_χ = Total number of customer segments $t_{c,P}$ = Index of all transactions including a product of type P made by customer c (sometimes represented without subscripts to ease reading) $T_{c,P}$ =Total number of transactions including a product of type P made by customer c
a)	Objective	Reveal segments of products that are related in terms of the strength with which they attract different customer-types	Reveal segments of customers that are related in terms of the strength with which they prefer different product-types
b)	Impose the aggregation constraint on	Customers. Define customer-types C = 1,2,…N_C, such that every individual customer, c, is of one, and only one, customer-type	Products. Define product-types P = 1,2,…,N_P such that every individual product, p, is of one, and only one, product-type
c)	Directly identify	Product segments: Π = 1,2,…, N_Π	Customer segments: χ = 1,2,…,N_χ

d)	Model Assumptions	The strength with which product p attracts customer-type C, relative to the strength with which p attracts other customer-types: $a_{p,C,t} = \alpha_{p,C,t} + \varepsilon_{p,C,t}$	The strength with which customer c prefers product-type P, relative to the strength with which c prefers other product-types: $u_{c,P,t} = v_{c,P,t} + \varepsilon_{c,P,t}$
e)		For a given purchase of product p, the probability that the purchase was made by a customer of type C_0 is the probability that product p attracted customer-type C_0 more strongly than it attracted any other customer-type $prob[\alpha_{p,C_0,t} \geq \alpha_{p,C,t}, \forall C \neq C_0]$	For a given choice by customer c, the probability that the purchase was of product-type P_0 is the probability that customer c prefers product-type P_0 more strongly than he/she prefers any other product-type $P[v_{c,P_0,t} \geq v_{c,P,t}, \forall P \neq P_0]$
f)		$\varepsilon_{p,C,t}$ are iid Gumbel type II extreme value	$\varepsilon_{c,P,t}$ are iid Gumbel type II extreme value
g)	Model	Attraction Model: $prob(C_0, t \mid p) = \exp\{\alpha_{p,C_0,t}\} \Big/ \sum_{C=1}^{N_C} \exp\{\alpha_{p,C,t}\}$ We refer to these probabilities as product p's customer mix since, all else equal, $prob(C_0, t \mid p)$ is the expected proportion of product p's customers who are customer-type C_0	Choice Model: $prob(P_0, t \mid c) = \exp\{v_{c,P_0,t}\} \Big/ \sum_{P=1}^{N_P} \exp\{v_{c,P,t}\}$ We refer to these probabilities as customer c's product choice shares since, all else equal, $prob(P_0, t \mid c)$ is the expected proportion of customer c's purchases that are product-type P_0
h)		$\alpha_{p,C,t} = \sum_{m=1}^{M} w_m x_{p,C,t}$ where w_m = attraction weight of characteristic m and $x_{p,C,t}$ = observed value of characteristic m for customer-type C and product p on the t^{th} transaction	$v_{c,P,t} = \sum_{j=1}^{J} z_j y_{c,P,t}$ where z_j = preference weight of characteristic j and $y_{c,P,t}$ = observed value of characteristic j for customer c and product-type P on the t^{th} transaction
i)	Latent class analysis	Infer product segments by identifying products that have similar customer mixes.	Infer customer segments by identifying customers that have similar product choice shares.

		$prob(C,t \mid p\varepsilon\Pi) = Q_\Pi * prob(C,t \mid p)$ where $Q_\Pi = \exp\{\theta_\Pi\} \Big/ \sum_{\Pi=1}^{N_\Pi} \exp\{\theta_\Pi\}$ is the relative size of each product segment in terms of the unconditional probability that a given product p is included in product segment Π, and θ_Π is the estimated product segment size parameter	$prob(P,t \mid c\varepsilon\chi) = R_\chi * prob(P,t \mid c)$ where $R_\chi = \exp\{\gamma_\chi\} \Big/ \sum_{\chi=1}^{N_\chi} \exp\{\gamma_\chi\}$ is the relative size of each customer segment in terms of the unconditional probability that a given customer c is included in customer segment χ, and γ_χ is the estimated customer segment size parameter
j)	Likelihood function	Letting H_p be the collection of all transactions in which product p was chosen, L(H_p) = $\sum_{\Pi=1}^{N_\Pi} Q_\Pi * L(H_p \mid \Pi)$ $= \sum_{\Pi=1}^{N_\Pi} [Q_\Pi * (\prod_{C=1}^{N_C} \prod_{t_{p,C}=1}^{T_{p,C}} prob(C, t_{p,C} \mid p\varepsilon\Pi))]$	Letting H_c be the collection of all transactions made by customers c, L(H_c) = $\sum_{\chi=1}^{N\chi} R_\chi * L(H_c \mid \chi)$ $= \sum_{\chi=1}^{N\chi} [R_\chi * (\prod_{P=1}^{N_p} \prod_{t_{c,P}=1}^{T_{c,P}} prob(P, t_{c,P} \mid c\varepsilon\chi))]$
k)	Results of MLE	Relative size of each product segment, Q_π and w_m^Π = attraction weight of characteristic m for products in product segment Π	Relative size of each customer segment, R_χ and z_j^χ = preference weight of characteristic j for customers in customer segment χ
l)	Bayesian calculation for set assignment	$prob(p\varepsilon\Pi \mid H_p) = L(H_p \mid \Pi) * Q_\Pi \Big/ \sum_{\Pi=1}^{N_\Pi} [L(H_p \mid \Pi) * Q_\Pi]$	$prob(c\varepsilon\chi \mid H_c) = L(H_c \mid \chi) * R_\chi \Big/ \sum_{\chi=1}^{N\chi} [L(H_c \mid \chi) * R_\chi]$

APPENDIX B
CONDITIONAL PROBABILITIES USED TO COMPARE RECOMMENDATIONS

	Product Segmentation Approach	Customer Segmentation Approach
Test 1: Objective of test	Recommend products in product segment Π for a customer of type C_0 to purchase	Recommend products of type P for a customer for a customer of type C_0 to purchase
Quantities estimated with approach	$prob(C_0\|\Pi)$ = probability, given a product in segment Π was chosen, the choice was made by a customer of type C_0 Q_Π = size of product segment Π ; i.e., unconditional probability product p is included in product segment Π	$prob(P \mid \chi)$ = probability, given the choice was made by a customer in segment χ , a product of type P was chosen
Quantities observed from sample	$prob(C_0)$ = proportion of customers that are of type C_0, i.e., unconditional probability customer c is of customer type C_0	$prob(\chi \mid C_0)$ = probability a customer of type C_0 is a member of customer segment χ[a]
Conditional probabilities on which recommendations are based	$prob(\Pi \mid C_0) = prob(C_0 \mid \Pi) * Q_\Pi / prob(C_0)$ = probability, given a customer of type C_0 did the choosing, a product from segment Π was chosen	$prob(P \mid C_0) = prob(P \mid \chi) * prob(\chi \mid C_0)$ = probability, given a customer of type C_0 did the choosing, a product of type P was chosen

Test 2: Objective of test	Recommend customers of type C for a product of product type P_0 to target	Recommend customers from customer segment χ for a product of product type P_0 to target
Quantities estimated with model	$prob(C \mid \Pi)$ = probability, given a product in segment Π was chosen, the choice was made by a customer of type C	$prob(P_0 \mid \chi)$ = probability, given a customer from customer segment χ did the choosing, a product of type P_0 was chosen R_χ = size of customer segment χ, i.e., unconditional probability customer c is included in customer segment χ
Quantities observed from sample	prob(Π $\mid P_0$) = probability a product of type P_0 is a member of product segment Π [b]	prob(P_0) = proportion of all products that are of type P_0, i.e., unconditional probability product p is of type P_0
Conditional probabilities on which recommendations are based	$prob(C \mid P_0) = prob(C \mid \Pi)\, prob(\Pi \mid P_0)$ = probability, given a product of type P_0 was chosen, a customer of type C made the choice.	$prob(\chi \mid P_0) = prob(P_0 \mid \chi) * R\chi / prob(P_0)$ = probability, given a product of type P_0 was chosen, a customer from customer segment χ made the choice.

[a] This is the method we used to estimate prob($\chi \mid C_0$), however, since the model gives no guidance on how to calculate this, other methods could be used

b This is the method we used to estimate prob($\Pi \mid P_0$), however, since the model gives no guidance on how to calculate this, other methods could be used

APPENDIX C
ADDITIONAL RULES USED TO COMPARE RECOMMENDATIONS

Rule for Recommendations	Product Segmentation Approach	Customer Segmentation Approach
	Test 1:	
Recommendation	Recommend products in product segment Π to a customer of type C_0 if:	Recommend products of product-type P to a customer of type C_0 if:
Fixed threshold	$prob(\Pi \mid C_0) > .50$	$prob(P \mid C_0) > .50$
Relative threshold	Π has highest $prob(\Pi \mid C_0)$	P has highest $prob(P \mid C_0)$
Better than chance I:	$prob(\Pi \mid C_0) >$ [1 / number of product segments]	$prob(P \mid C_0) >$[1 / number of product-types]
Better than chance II:	$prob(\Pi \mid C_0) >$ [number of products in product segment Π / total number of products]	$prob(P \mid C_0) >$ [number of products of type P / total number of products]
	Test 2:	
Recommendation	Recommend that a product of type P_0 target customers of customer-type C if:	Recommend that a product of type P_0 target customers in customer segment χ if:
Fixed threshold:	$prob(C \mid P_0) > .50$	$prob(\chi \mid P_0) > .50$
Relative threshold:	C with highest $prob(C \mid P_0)$	χ with highest $prob(\chi \mid P_0)$
Better than chance I:	$prob(C \mid P_0) >$ [1 / number of customer-types]	$prob(\chi \mid P_0) >$ [1 / number of customer segments]
Better than chance II:	$prob(C \mid P_0) >$ [number of customers of type C / total number of customers	$prob(\chi \mid P_0) >$ [number of customers in customer segment χ/ total number of customers]

APPENDIX D
COMPARISON OF MEAN HIT RATES FOR ADDITIONAL RECOMMENDATION RULES

	PS Mean Hit Rates	CS Mean Hit Rates	Comparison of PS Mean Hit Rate & CS Mean Hit Rate (p-value)
Test 1:			
For withheld student of type C_0, what percent of the courses actually taken by the student were recommended by the approach			
Recommend if probability of product segment (PS) or product-type (CS) is:			
Greater than .50	.15	.00	< .01
The highest	.51	.39	< .01
Greater than [1/no. of sets of courses]	.69	.46	< .01
Greater than [no. of courses in set / total no. of courses]	.56	.31	< .01
Test 2:			
For withheld course of type P_0, what percent of the students who took the course were identified as being in the target segment			
Recommend if probability of customer-type (PS) or customer segment (CS) is:			
Greater than .50	.00	.45	< .01
The highest	.23	.40	.02
Greater than [1 / no. of sets of students]	.41	.45	.41
Greater than [no. of students in set / total no. of students]	.30	.43	.05

APPENDIX E
PRODUCT SEGMENTATION AND CUSTOMER SEGMENTATION APPROACH SPECIFICATIONS

PS Approach	Deterministic Component of Attraction Weight for Product *p*	Examples of Student-type Attraction Weights
Model 1	$a^p_{C,t} = w_1 D_{inv\ bank,t} + w_2 D_{corp\ fin,t} + w_3 D_{tech\ mgr,t} + w_4 D_{gen\ mgr,t} + w_5 D_{product\ mgr,t} + w_6 D_{cons,t}$ where $D_{C,t}$ is a dummy variable that takes on a value equal to 1 if the customer making the t^{th} transaction is of customer-type *C*, (i.e., took a job of type *C*) and takes on a value of 0 otherwise, and where w_m is the attraction weight for characteristic *m*	$a^p_{inv.bank,t} = w_1$ $a^p_{corp.fin,t} = w_2$
Model 2	$a^p_{C,t} = w_1 D_{inv\ bank,t} + w_2 D_{corp\ fin,t} + w_3 D_{tech\ mg,tr} + w_4 D_{gen\ mgr,t} + w_5 D_{product\ mgr,t} + w_6 D_{cons,t} + w_7 D_{tech\ degree,t}$ where $D_{tech\ degree,t}$ is a dummy variable that takes on a value of 1 if the customer making the t^{th} transaction has a technical undergraduate degree and takes on a value of 0 otherwise	$a^p_{inv.bank,t} = w_1 + w_7 D_{tech\ degree,t}$ $a^p_{corp.fin,t} = w_2 + w_7 D_{tech\ degree,t}$
Model 3	$a^p_{C,t} = w_1 D_{inv\ bank,t} + w_2 D_{corp\ fin,t} + w_3 D_{tech\ mgr,t} + w_4 D_{gen\ mgr,t} + w_5 D_{product\ mgr,t} + w_6 D_{cons,t} + w_7 D_{tech\ degree,t} + w_8 X_{eval,t}*D_{inv\ bank,t} + w_9 X_{eval,t}*D_{corp\ fin,t} + w_{10} X_{eval,t}*D_{tech\ mgr,t} + w_{11} X_{eval,t}*D_{gen\ mgr,t} + w_{12} X_{eval,t}*D_{product\ mgr,t} + w_{13} X_{eval,t}*D_{cons,t}$ where $X_{eval,t}$ is a continuous variable representing the average evaluation score for the product chosen on the t^{th} transaction, normalized on a scale ranging from 1 to 5	$a^p_{inv.bank,t} = w_1+w_7 D_{tech\ degree,t}+w_8 X_{eval,t}$ $a^p_{corp.fin,t} = w_2+w_7 D_{tech\ degree,t}+w_9 X_{eval,t}$

CS Approach	Deterministic Component of Preference Weight for Customer *c*	Examples of Course-type Preference Weights
Model 4	$v^c_{P,t} = z_1 D_{FIN,t} + z_2 D_{MIS,t} + z_3 D_{MAN,t} + z_4 D_{MKT,t}$ where $D_{P,t}$ is a dummy variable that takes on a value equal to 1 if the product chosen on the t^{th} transaction is of product-type P and takes on a value of 0 otherwise, and where z_j is the importance weight for characteristic j	$v^c_{FIN,t} = z_1$ $v^c_{MIS,t} = z_2$
Model 5	$v^c_{P,t} = z_1 D_{FIN,t} + z_2 D_{MIS,t} + z_3 D_{MAN,t} + z_4 D_{MKT,t} + z_5 X_{eval,t}$	$v^c_{FIN,t} = z_1 + z_5 X_{eval,t}$ $v^c_{MIS,t} = z_2 + z_5 X_{eval,t}$
Model 6	$v^c_{P,t} = z_1 D_{FIN,t} + z_2 D_{MIS,t} + z_3 D_{MAN,t} + z_4 D_{MKT,t} + z_5 X_{eval,t} + z_6 D_{FIN,t} * D_{tech\ degree,t}$ $+ z_7 D_{MIS,t} * D_{tech\ degree,t} + z_8 D_{MAN,t} * D_{tech\ degree,t} + z_9 D_{MKT,t} * D_{tech\ degree,t}$	$v^c_{FIN,t} =$ $z_1 + z_5 X_{eval,t} + z_6 D_{tech\ degree,t}$ $v^c_{MIS,t} =$ $z_2 + z_5 X_{eval,t} + z_7 D_{tech\ degree,t}$

APPENDIX F
PARAMETER STABILITY ACROSS N-FOLD BOOTSTRAP ESTIMATION

	Product Segmentation Model 1 4-Product Segment Solution				Customer Segmentation Model 5 2-Customer Segment Solution		
Test 1	The mean and the standard deviation[a] of the value for each parameter, estimating model 326 times, holding out one student each time.						
Customer-Types	**Product Seg 1**	**Product Seg 2**	**Product Seg 3**	**Product Seg 4**	**Product-Types**	**Customer Seg 1**	**Customer Seg 2**
Inv Banker	.93	-1.94	-.95	-.10	FIN	.88	-.09
	(.01)	(.05)	(.05)	(.03)		(.01)	(.01)
Corp Finance	.25	-.99	-.98	-.57	MIS	-.56	.67
	(.01)	(.01)	(.04)	(.01)		(.05)	(.05)
IT Manager	-1.34	.61	-.95	-.69	MAN	-.06	.56
	(.02)	(.01)	(.02)	(.03)		(.02)	(.02)
General Mgr	-.61	-.51	-.08	-.15	MKT	-.09	.33
	(.01)	(.01)	(.02)	(.01)		(.03)	(.03)
Product Mgr	-.99	-.36	.48	.07	Course	4.53	3.52
	(.02)	(.01)	(.04)	(.01)	Eval	(.00)	(.00)
Consultant	.06	.44	.22	.17			
	(.01)	(.01)	(.01)	(.01)			
Product	.29	.16	.17	.38	Customer	.46	.54
Seg Size	(.02)	(.02)	(.03)	(.03)	Seg Size	(.00)	(.01)

	Product Segmentation Model 1 4-Product Segment Solution				Customer Segmentation Model 5 2-Customer Segment Solution		
Test 2	The mean and the standard deviation[a] of the value for each parameter, estimating model 32 times, holding out one course each time.						
Customer-Types	**Product Seg 1**	**Product Seg 2**	**Product Seg 3**	**Product Seg 4**	**Product-Types**	**Customer Seg 1**	**Customer Seg 2**
Inv Banker	.92	-1.88	-1.40	-.19	FIN	.99	-.12
	(.01)	(.08)	(.18)	(.01)		(.03)	(.08)
Corp Finance	.23	-.99	-1.18	-.60	MIS	-.57	.66
	(.01)	(.03)	(.21)	(.01)		(.04)	(.04)
IT Manager	-1.37	.63	-.91	-.65	MAN	-.04	.52
	(.05)	(.04)	(.14)	(.01)		(.03)	(.04)
General Mgr	-.64	-.52	-.19	-.15	MKT	-.01	.34
	(.01)	(.01)	(.09)	(.01)		(.04)	(.05)
Product Mgr	-1.05	-.39	.34	.10	Course	4.57	3.50
	(.02)	(.01)	(.27)	(.01)	Eval	(.02)	(.02)
Consultant	.04	.45	-.02	.20			
	(.01)	(.01)	(.25)	(.01)			
Product	.28	.16	.13	.43	Customer	.50	.50
Seg Size	(.04)	(.06)	(.10)	(.08)	Seg Size	(.03)	(.00)

[a] Standard deviation in parentheses
Note: variation in estimated parameters for Product Segment 3 in Test 2 did not impact product groupings in a way that changed calculated recommendation probabilities made

APPENDIX G
ATTRIBUTE-BASED CO-CLUSTERING ALGORITHM

Begin with a random co-clustering (ρ, γ)

Repeat until convergence:

Step 1. Update co-cluster prototypes

    For g=1:k

      For h=1:l

        Compute the mean vector $[\mu_{gh}, \mathbf{C\mu_{gh}}^T, \mathbf{P\mu_{gh}}^T]^T$ of all the customer-product pairs in co-cluster g-h

      End for

    End for

Step 2a. Update ρ - assign each row to the row cluster that minimizes the error

Step 2b. Update γ - assign each column to the column cluster that minimizes the error

Return (ρ, γ)

Step 1 minimizes the objective function due to the property of the mean, while steps 2 and 3 directly minimize the objective function. The objective function hence decreases at every iteration and is bounded from below by zero, and the algorithm is guaranteed to converge to a local minima.

BIBLIOGRAPHY

Ansari, Asim, Skander Essegaier, and Rajeev Kohli (2000), "Internet Recommendation Systems," *Journal of Marketing Research*, 37 (August), 363-375.

Ariely, Dan, John G. Lynch, Jr., and Manuel Aparicio IV (2004), "Learning by Collaborative and Individual-Based Recommendation Agents," *Journal of Consumer Psychology*, 14 (1 & 2), 81-95.

Banerjee, Arindam, Inderjit Dhillon, Joydeep Ghosh, Srujana Merugu, and Dharmendra S. Modha (forthcoming), "A Generalized Maximum Entropy Approach to Bregman Proposed attribute-based co-clustering and Matrix Approximation," *Journal of Machine Learning Research*.

Breese, Jack, David Heckerman, and Carl Kadie (1998), "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," in *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*. Madison, WI: Morgan Kauffmann Publisher, 43-52.

Canny, John (2002), "Collaborative Filtering With Privacy," paper presented at the *IEEE Symposium on Security and Privacy*, Oakland, CA (May 2002).

Cheng, Yizong and George M. Church (2000), "Biclustering of Expression Data," in *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology*. AAAI Press, 93-103.

Chien, Yung-Hsin and Edward I. George (1999), "A Bayesian Model for Collaborative Filtering," working paper, IROM Department, University of Texas at Austin.

Cho, Hyuk, Inderjit S. Dhillon, Yuqiang Guan, and Suvrit Sra (2004), "Minimum Sum-squared Residue Co-clustering of Gene Expression Data," in Proceedings of the Fourth SIAM Data Mining Conference. Lake Buena Vista, FL: 22-24.

Convenience Store News (2004), "New Service Tracks Trends in Packaged Goods," 40 (12), 26.

Cooper, Lee G. (1988), "Competitive Maps: The Structure Underlying Asymmetric Cross Elasticities," *Management Science*, 34 (6), 707-723.

Daniels, Alex (2004), "Wal-Mart Exec Says Stores Too Cluttered," *Arkansas Democrat-Gazette*, October 20.

DeSarbo, Wayne S. and Geert De Soete (1984), "On the Use of Hierarchical Clustering for the Analysis of Nonsymmetric Proximities," *Journal of Consumer Research*, 11 (1), 601-610.

Dhillon, Inderjit S., Subramanyam Mallela, and Dharmendra S. Modha (2003), "Information-Theoretic Co-clustering," in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge, Discovery and Data Mining*. New York, NY: ACM Press, 89-98.

Economist (2005), "Extinction of the Predator," 376 (8433), 63-65.

Elrod, Terry and Michael P. Keane (1995), "A Factor-Analytic Probit Model for Repressing the Market Structure in Panel Data," *Journal of Marketing Research*, 32 (1), 1-16.

Erdem, Tulin (1996), "A Dynamic Analysis of Market Structure Based on Panel Data," *Marketing Science*, 15 (4), 359-378.

----- and Russell S. Winer (1999), "Econometric Modeling of Competition: A Multi-Category Choice-Based Mapping Approach," *Journal of Econometrics*, 89 (1/2), 159-175.

Fraser, Cynthia and John W. Bradford (1983), "Competitive Market Structure Analysis: Principal Partitioning of Revealed Substitutabilities," *Journal of Consumer Research*, 10 (1), 15-30.

George, Thomas and Srujana Merugu (2005), "A Scalable Collaborative Filtering Framework Based on Proposed attribute-based co-clustering," in *Proceedings of the Fifth IEEE International Conference on Data Mining*. Houston, TX: 625-628.

Godfrey, Andrea, Leigh McAlister and Maytal Saar-Tsechansky (2007), "Identifying Customer-Centric, Cross-Category Product Groups: A Product Segmentation Approach and its Relationship to Customer Segmentation Approaches," working paper, Marketing Department, The University of Texas at Austin.

Goldberg, David, David Nichols, Brian M. Oki, and Douglas Terry (1992), "Using Collaborative Filtering to Weave an Information Tapestry," Communications of the ACM, 35 (12), 61-70.

Grover, Rajiv and William R. Dillon (1985), "A Probabilistic Model for Testing Hypothesized Hierarchical Market Structures," *Marketing Science*, 4 (4), 312-335.

----- and V. Srinivasan (1987), "A Simultaneous Approach to Market Segmentation and Market Structuring," *Journal of Marketing Research*, 24 (2), 139-153.

Haubl, Gerald and Valerie Trifts (2000), "Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids," *Marketing Science*, 19 (1), 4-21.

Heilman, Carrie M. and Douglas Bowman (2002), "Segmenting Consumers Using Multiple-Category Purchase Data," *International Journal of Research in Marketing*, 19 (3), 225-252.

Herlocker, Jonathan L., Joseph A. Konstan, Al Borchers, and John Riedl (1999), "An Algorithmic Framework for Performing Collaborative Filtering," in *Proceedings of the Twenty-second International Conference Research and Development in Information Retrieval*. New York, NY: M.A. Hearst, F.F. Gey, and R. Tong, Eds., 230-237.

-----, -----, Loren G. Terveen, and John T. Riedl (2004), "Evaluating Collaborative Filtering Recommender Systems," *ACM Transactions on Information Systems*, 22 (1), 5-53.

Hoffmann, Thomas ad Jan Puzicha (1999), "Latent Class Models for Collaborative Filtering," in *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*. Madison, WI: Morgan Kauffmann Publisher, 688-693.

Kamakura, Wagner A. and Gary J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (4), 379-390.

Konstan, Joseph A., Bradley N. Miller, David Maltz, Jonathon L. Herlocker, Lee R Gordon, and John T. Riedl (1997), "GroupLens: Applying Collaborative Filtering to Usenet News," *Communications of the ACM*, 40 (3), 77-87.

Levy, Michael and Barton A. Weitz (2006), *Retailing Management*, 6th ed. Boston: McGraw-Hill / Irwin.

McAlister, Leigh, Edward I. George, and Yung-Hsin Chien (2007), "A Basket-Mix Model to Identify Products that Draw Cherry-Pickers and Products that are Trip-Generators," working paper, Marketing Department, The University of Texas at Austin.

McTaggert, Jenny (2006), "Wal-Mart Unveils New Segmentation Scheme," *Progressive Grocer*, 85 (14), 10-12.

Mild, Andreas and Martin Natter (2002), "Collaborative Filtering or Regression Models for Internet Recommendation Systems?" *Journal of Targeting, Measurement and Analysis for Marketing*, 10 (4), 304-313.

Mild, Andreas and Thomas Reutterer (2003), "An Improved Collaborative Filtering Approach for Predicting Cross-Category Purchases Based on Binary Market Basket Data," *Journal of Retailing and Consumer Services*, 10 (2003), 123-133.

Mitchell, Thomas M. (1997), *Machine Learning.* Boston: McGraw Hill.

Murthi, B. P. S. and Sumit Sarkar (2003), "The Role of the Management Sciences in Research on Personalization," *Management Science*, 49 (10), 1344-1362.

Ramaswamy, Venkatram and Wayne S. DeSarbo (1990), "SCULPTRE: A New Methodology for Deriving and Analyzing Hierarchical Product-Market Structures From Panel Data," *Journal of Marketing Research*, 27 (4), 418-427.

Rao, Vithala R. and Darius Sabavala (1981), "Inference of Hierarchical Choice Processes from Panel Data," *Journal of Consumer Research*, 8 (1), 85-96.

Russell, Gary J. and Ruth N. Bolton (1988), "Implications of Market Structure for Elasticity Structure," *Journal of Marketing Research*, 25 (3), 229-241.

Shugan, Steven M. (1987), "Estimating Brand Positioning Maps Using Supermarket Scanning Data," *Journal of Marketing Research*, 24 (1), 1-18.

Wedel, Michel and Jan-Benedict E. M. Steenkamp (1991), "A Clusterwise Regression Method for Simultaneous Fuzzy Market Structuring and Benefit Segmentation," *Journal of Marketing Research*, 27 (November), 385-396.

West, Patricia M., Dan Ariely, Steve Bellman, Eric Bradlow, Joel Huber, Eric Johnson, Barbara Kahn, John Little, and David Schkade (1999), "Agents to the Rescue?" *Marketing Letters*, 10 (3), 285-300.

Ying, Yuanping, Fred Feinberg, and Michel Wedel (2006), "Leveraging Missing Ratings to Improve Online Recommendation Systems," *Journal of Marketing Research*, 43 (August), 355-365.

VITA

Andrea Lynn Godfrey was born in Saint John, New Brunswick, Canada on November 11, 1973, the daughter of Beverly Irene Godfrey and Grant William Godfrey. After completing her work at Saint John High School in 1991, she entered McGill University in Montreal, Quebec, Canada where she received the degree of Bachelor of Engineering (Civil) in June 1996. During the following four years, she was employed as an Engineer-In-Training at Godfrey Associates Ltd. and became a registered Professional Engineer in June 2000. In September 2000, she entered Babson College in Wellesley, Massachusetts where she received the degree of Master of Business Administration in May 2002. In September 2002, she entered the Graduate School of The University of Texas at Austin.

Permanent Address: 6805 Wood Hollow Drive, Apt. 362, Austin, Texas 78731

This dissertation was typed by the author.